UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

CENTRO DE BIOTECNOLOGIA

PROGRAMA DE PÓS-GRADUAÇÃO EM BIOLOGIA CELULAR E MOLECULAR

Michèle Claire Breton

ESTUDO DO TRANSCRIPTOMA EM FLORES DE *EUCALYPTUS GRANDIS*

Orientadores: **Prof. Dr. Jeverson Frazzon**

Prof. Dr. Giancarlo Pasquali

Porto Alegre, junho de 2011

Michèle Claire Breton

Bióloga pela Universidade Estadual de Londrina-PR
Especialista em Genética pela Universidade Estadual de Maringá-PR e Embrapa Soja-PR
Mestre em Genética e Melhoramento pela Universidade Estadual de Londrina e Embrapa Soja-PR

ESTUDO DO TRANSCRIPTOMA DE FLORES DE *EUCALYPTUS GRANDIS*

Tese apresentada junto ao Programa de Pós-graduação em Biologia Celular e Molecular do Centro de Biotecnologia da Universidade Federal do Rio Grande do Sul como requisito para a obtenção do grau de Doutor.

Orientadores: **Prof. Dr. Jeverson Frazzon**
Prof. Dr. Giancarlo Pasquali

Porto Alegre, junho de2011

BANCA EXAMINADORA

_____

Prof. Dr. Jeverson Frazzon – *Orientador*
Instituto da Ciência e Tecnologia de Alimentos, UFRGS

_____

Prof. Dr. Giancarlo Pasquali – *Orientador*
Instituto de Biociências e Centro de Biotecnologia, UFGRS

_____

Prof. Dr. Gilberto Sachetto-Martins
Instituto de Biologia, Universidade Federal do Rio de Janeiro

_____

Prof. Dr. Charley C. Staats
Instituto de Biociências e Centro de Biotencologia, UFRGS

_____

Dr. Raul Antonio Sperotto
Bolsista de Pós-doutorado PRODOC - Capes
Programa de Pós-graduação em Botânica - UFRGS

Porto Alegre, 17 de junho de 2011

Este trabalho foi desenvolvido, em sua totalidade, no Laboratório de Biologia Molecular Vegetal do Centro de Biotecnologia da Universidade Federal do Rio Grande do Sul. O material biológico utilizado neste trabalho foi cedido pela Aracruz Celulose, Unidade de Guaíba, RS. Os auxílios financeiros foram provenientes da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Ministério da Educação, e do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Ministério da Ciência e Tecnologia.

Ao meu pai *Jean-Marie* e minha mãe *Rosa*;

À minha irmã *Nicole*;

aos meus sobrinhos *Denis* e *Alan*;

e a vida, por ter feito da menina ingênua uma mulher forte...

dedico.

AGRADECIMENTOS INSTITUCIONAIS

Aos Prof. Dr. *Jeverson Frazzon* e Prof. Dr. *Giancarlo Pasquali*, pela oportunidade e orientação ao longo deste trabalho;

Aos Engenheiros Florestais *Aurélio Mendes Aguiar* e *João Bauer*, por nos ceder, muito atenciosamente, material biológico para as pesquisas em nosso laboratório;

Ao Dr. *Sandro da Silva Camargo*, pelos ensinamentos e parceria nos trabalhos de bioinformática;

Ao Prof. Dr. *Rogério Margis* e sua equipe, principalmente à *Andreia T. Zollet*, pelo auxílio na construção das bibliotecas e cálculos de Real Time PCR;

À Profa. Dra. *Maria Luiza Saraiva Pereira* e *Marcia Maria A. Naschenveng Pinheiro Margis*, pela cessão do equipamento de Real Time PCR;

Aos Prof. Dr. *Rogério Margis*, Prof. Dr *Arnaldo Zaha* e Prof. Dr. *Arthur Germano Fett Neto*, membros da comissão de acompanhamento do doutorado;

À *Silvia Centeno* e *Luciano Saucedo*, pela eficiência e carinho com que acolhem os alunos do programa;

Aos Técnicos de laboratório *Milton* e *Paulo*, pelo trabalho dedicado em fornecer materiais para realização dos experimentos;

Aos docentes integrantes do Programa de Pós-graduação em Biologia Celular e Molecular, pelos ensinamentos e formação profissional;

À CAPES e CNPq, pelo apoio financeiro;

AGRADECIMENTOS PESSOAIS

Aos meus pais *Jean-Marie Breton* e *Rosa Chagas Breton*, por toda minha vida: amor incondicional, carinho, cobrança, formação, apoio, responsabilidade, caráter e acima de tudo, pelas lições que me fizeram ter um bom coração;

À minha irmã *Nicole Françoise Breton*, com carinho, pelo amor, incentivo, apoio, paciência em me ouvir, e infindáveis risadas;

Aos meus sobrinhos *Denis Breton de Carvalho* e *Alan Breton de Carvalho*, por ser minha inspiração diária e razão da minha caminhada, me surpreender a cada dia, e me incentivar a ser uma pessoa melhor;

Ao meu querido amigo *José Fernando de Lacerda Machado Junior*, pela eterna amizade, confiança, carinho e cumplicidade, sem os quais eu não saberia viver;

Aos meus queridos amigos conselheiros *Sandro da Silva Camargo* e *Hugo da Silva Pinto*, pelas inúmeras vezes que trouxeram os pés desta sonhadora de volta a Terra, por terem ouvido meus lamentos, e terem sempre acreditado que eu ia vencer;

À minha amiga *Luciana Foss*, pelas conversas, fofocas, tardes divertidas na UFRGS e pela valiosa amizade;

Ao meu querido amigo *William Nunes da Silva Junior*, pela infinita amizade desde os tempos de faculdade, que, apesar das intempéries do tempo, resistiu e ressurgiu no momento em que eu mais precisava;

À minha amiga *Fernanda Macedo Bastolla*, pela amizade dedicada e pelas inúmeras discussões sobre "Harry Potter" quando precisávamos fugir da realidade;

Às minhas queridas amigas, *Luisa Abruzzi de Oliveira, Marta Dalpian Heis* e *Rochele Kirch*, pelas confidencias, carinho, paciência e divertidíssimas "indiadas";

Ao meu querido amigo *Guilherme Pizzoli* pelo carinho, pela companhia e pelas filosofias sobre as bactérias;

Ao meu amigo *Felipe dos Santos Maraschin*, cuja convivência fez renascer o melhor de mim;

Aos meus amigos e mestres *Jeverson Frazzon* e *Giancarlo Pasquali*, pelos ensinamentos, amizade e exemplo profissional;

Às minhas amigas *Gabriela M. V. V. Bandeira, Fernanda Sperb, Marina Tagliaro Jahns* e *Sinara Ártico*, pela recepção no Laboratório, pela maravilhosa convivência e amizade;

Aos meus amigos "Iron-Sulfur Clusters" *Christine Bierhals, Gustavo Pelicioli Riboldi, Eduardo Preusser de Mattos* e *Leonardo Pedrazza*, pelo companheirismo, músicas e momentos inesquecíveis no laboratório;

A nova geração de amigos do LBMV: *Genaro Azambuja Athaydes, Juliana Dametto Kras Borges, Camila Bedin Scalco, Gabriel S. D´Almeida, Patrícia Picolotto, Ricardo Luís Mayer Weber* e *Beatriz Wiebke Strohm,* pela convivência, companheirismo, muitas risadas, festas divertidas e inúmeros cafés;

Às Dra*. Ines Guimarães da Silveira*, médica reumatologista, e a Dra. *Vanessa Moreira*, psicóloga, por me manterem saudável, física e psicologicamente, durante o difícil período do doutorado;

À todos os amigos que fiz ao longo da vida, pelos caminhos que trilhei, sintam-se homenageados neste parágrafo, pois trago dentro de meu coração um pouquinho de cada um de vocês, obrigada pela valorosa amizade!

À vida: muitas vezes dura e difícil, para nos tornar fortes; muitas vezes triste e desestimulante, para testar nossa capacidade de recomeço e fé; e poucas vezes linda e apaixonante, para nos lembrar que por estes momentos preciosos, vale a pena viver cada segundo!

À Deus, por tudo!

”Non... rien de rien...	*“Não... nada...*
Non... je ne regrette rien	*Não... eu não lamento nada*
Ni le bien qu´on ma fait,	*Nem o bem que me fizeram*
Ni le mal – tout ça m´est bien égal!	*Nem o mal – isso tudo tanto faz!*
Balayé les amours	*Varridos os amores*
Avec leurs tremolos	*E todos os seus temores*
Balayé pour toujours	*Varridos para sempre*
Je repars à zero”	*Recomeço do zero”*

Versos da canção *Non, Je Ne Regrette Rien* – Edith Piaf

ÍNDICE

RELAÇÃO DE ABREVIATURAS E SÍMBOLOS

Kg	Quilograma(s)
Cia	Companhia
m^3	Metro(s) cúbico(s)
Há	Hectare(s)
GM	Geneticamente Modificado
Mb	Megabases
2n	Número diplóide de cromossomos de uma espécie
M	Metro(s)
Mm	Milímetro(s)
Cm	Centímetro(s)
°C	Grau(s) Celsius
°S	Graus sul
RT-qPCR	Reação em cadeia da DNA polimerase em tempo real ou quantitativa precedida de transcrição reversa
DNA	Ácido desoxirribonucléico
RNA	Ácido ribonucléico
cDNA	Ácido desoxirribonucléico complementar
EST	Marcas ou etiquetas de sequências expressas - do inglês, *Expressed Sequence Tags*;
Pb	Par(es) de bases

RESUMO

BRETON, Michèle Claire, D.Sc., Universidade Federal do Rio Grande do Sul, junho de 2011. **Estudo do transcriptoma em flores de *Eucalyptus grandis*.** Orientadores: Jeverson Frazzon e Giancarlo Pasquali.

A indústria de base florestal é estratégica para o Brasil devido ao seu perfil fortemente exportador. O setor responde pela segunda posição na balança comercial do agronegócio brasileiro, ficando atrás somente da soja em grão. Atualmente, a área ocupada com florestas de eucaliptos no Brasil atinge 1,9 milhões de ha. Diante da importância sócio-econômica que a silvicultura desempenha no mercado brasileiro e do aumento progressivo das áreas plantadas com florestas de *Eucalyptus*, o grande desafio para o melhoramento do eucalipto está na integração da biotecnologia mais avançada ao seu cultivo, o que compreende a identificação de genes controladores das características de importância econômica e ambiental e a transferência destes genes entre árvores por meio de cruzamentos controlados ou modificação direcionada. Portanto, os objetivos deste estudo foram apresentados em 3 capítulos distintos: I - a identificação de genes expressos em flores de *E. grandis* em processo de antese; II - o estudo mais refinado de mineração e identificação de genes potencialmente codificadores de fatores de transcrição presentes no genoma de *E. grandis*; e III - seleção de 50 genes cujas expressões mostraram-se constitutivas entre folhas e xilemas de *E. grandis* e xilema de *E. globulus*, pela técnica de hibridização de microarranjos de DNA. No capítulo I é apresentada uma breve fundamentação teórica sobre as flores de *E. grandis* e a importância do estudo da expressão gênica e da identificação de genes envolvidos em determinados processos metabólicos e fisiológicos das plantas. Nos resultados, apresentados juntamente com a discussão, estão apresentados o conjunto de transcritos identificados e a anotação dos mesmos conforme as bibliotecas geradas. As sequências de genes expressos estão fundamentalmente envolvidas na manutenção do órgão, na senescência e em respostas a estímulos ambientais. Também são mostrados resultados obtidos por RT-qPCR para genes selecionados a partir das anotações cujo perfil de transcrição foi avaliado para as partes da flor, folha e xilema. Ao final do capítulo, é feita uma breve descrição da metodologia e das conclusões referente a este estudo. A partir dos dados anotados dos genes expressos nas bibliotecas de flores e botões florais descritos no primeiro capítulo, foram encontradas algumas famílias de fatores de transcrição, dentre elas, a família Dof, encontrada nas bibliotecas de carpelos/receptáculos florais. Assim, um segundo capítulo foi redigido na forma de manuscrito de artigo científico a ser submetido ao periódico *BMC Plant Biology*, em língua inglesa. Após a fundamentação teórica sobre os fatores Dof em plantas, foram apresentados os resultados obtidos de um estudo mais refinado de mineração e identificação de genes potencialmente codificadores destes fatores de transcrição presentes no genoma de *E. grandis*. Posteriormente, uma quantificação dos níveis de mRNA para alguns dos genes *Dof* de *E. grandis* a partir da técnica de RT-qPCR foi realizada. A análise foi realizada para órgãos diferentes da planta e em plântulas submetidas a estresses abióticos e sinalização por reguladores de crescimento vegetais. Os resultados e a discussão deste capítulo também são mostrados em uma única sessão, assim como a descrição da metodologia e as conclusões sobre esta etapa. O terceiro e último capítulo, também apresentado na forma de manuscrito de artigo científico a ser submetido ao periódico *BMC Plant Biology*, em língua inglesa, é composto de um estudo paralelo aos temas citados acima, realizado a partir da seleção de 50 genes cujas expressões mostraram-se constitutivas entre folhas e xilemas de *E. grandis* e xilema de *E. globulus*, pela técnica de hibridização de microarranjos de DNA. Dos 50 genes selecionados e anotados, oito foram selecionados para a validação por RT-qPCR em seis espécies de *Eucalyptus* e em três órgãos diferentes (flor, folha e xilema). As expressões destes genes candidatos foram comparadas àquelas de sete genes normalizadores tradicionais, usados frequentemente em estudos de avaliação da expressão gênica em plantas. Como nos demais capítulos, as conclusões referentes a este estudo, bem como a metodologia empregada, são apresentadas ao final do capítulo. Na parte final desta tese, constam as conclusões gerais sobre a totalidade do trabalho conduzido.

ABSTRACT

BRETON, Michèle Claire, D.Sc., Universidade Federal do Rio Grande do Sul, june, 2011. **Transcriptome study of *Eucalyptus grandis* flowers**. Advisors: Jeverson Frazzon and Giancarlo Pasquali.

The forest industry is strategic to Brazil due to its strong export profile. The sector accounts for the second position in the trade brazilian agribusiness balance. Currently, the area planted with *Eucalyptus* forests in Brazil reached 1.9 million/ha. Given the socioeconomic importance that forestry plays in the brazilian market and the gradual increase in areas planted with *Eucalyptus* forest, the great challenge for the *Eucalyptus* breeding is the integration of advanced biotechnology to cultivation, which includes the identification of economic importance genes controlling environmental effects and transfer of genes between trees through intersections controlled or directed modification. The objectives of this study were presented in three distinct chapters: I - the identification of expressed genes in *E. grandis* flowers in the anthesis process; II – more refined study, mining and identification of genes potentially encoding transcription factors present in the *E. grandis* genome and III - selection of 50 genes whose expressions were shown to be constitutive in *E. grandis* xylem and leaves and *E. globulus* xylem, by DNA microarray hybridization tecnique. In Chapter I, we present a brief theoretical background on the *E. grandis* flowers and the importance of studying gene expression and identifying genes involved in metabolic and physiological processes of plants. The results, presented transcripts sets identified and annotation the libraries generated. The sequences of expressed genes are mainly involved in organ maintaining, senescence and in responses to environmental stimuli. Also shown are results obtained by RT-qPCR for selected genes from the notes whose transcription profile was assessed for part of the flower, leaf and xylem. At the end of the chapter is a brief description of the methodology and conclusions about this study. From the data recorded in the libraries of genes expressed in flowers and flower buds described in the first chapter, we found some families of transcription factors, among them, the Dof family, found in the libraries of carpel/floral receptacles. Thus, a second chapter was drafted as a manuscript of a scientific paper to be submitted to the journal BMC Plant Biology. After the theoretical DOF on the factors in plants, we presented the results of a more refined mining and identification of genes potentially encoding these transcription factors present in the *E. grandis* genome. Subsequently, a quantification of mRNA levels for some of *E. grandis* Dof genes from the RT-qPCR was performed. The analysis was performed for different plant organs and in seedlings subjected to abiotic stress signaling and plant growth regulators. The results and discussion of this chapter are also shown in a single session, as well as a description of the methodology and conclusions about this step. The third and final chapter, also presented as a manuscript of a scientific paper to be submitted to the journal BMC Plant Biology, is composed of a parallel study to the subjects mentioned above, performed by the selection of 50 genes whose expression showed is constitutive of *E. grandis* leaves and xylem and *E.globulus* xylem, by the DNA microarray hybridization tecnique. Of the 50 selected genes and annoted, eight were selected for validation by RT-qPCR in six *Eucalyptus* species and three different organs (flower, leaf and xylem). The expression of these candidate genes were compared to those of traditional standard-seven genes, often used in evaluation studies of gene expression in plants. As in other chapters, the findings of this study and the methodology employed, are included at the end of the chapter. In the final part of this thesis, contains the general conclusions about the totality of the work conducted.

INTRODUÇÃO

4. A Origem do *Eucalyptus*

De ocorrência natural na Austrália, o gênero *Eucalyptus* pertence à família Myrtaceae e possui cerca de 600 espécies adaptadas a diversas condições de solo e clima, todas denominadas pelo termo popular "eucalipto". Desta grande variedade de espécies, apenas duas não são originárias da Austrália: *E. urophylla* e *E. deglupta*. A diversidade de espécies decorre, dentre outros, do hábito alógamo destas árvores, da resposta à pressão de seleção causada pelas alterações do meio ambiente e do próprio processo de deriva e especiação [1]. A maioria das espécies conhecidas de *Eucalyptus* é típica de florestas altas, cujas árvores atingem alturas que variam de 30 a 50 m, e de florestas abertas, com árvores menores, atingindo alturas entre 10 e 25 m. Cerca de 30 a 40 espécies são arbustivas [2, 3].

O gênero *Eucalyptus* possui um genoma de aproximadamente 630 Mb e a análise cariotípica de 108 espécies indicou que 95 apresentam um número mínimo básico de cromossomos 2n = 22, da mesma forma que a maioria dos gêneros da família Myrtaceae [4-6].

Acredita-se que durante o processo evolutivo, o *Eucalyptus* adaptou-se a condições de baixo conteúdo de nutrientes no solo. Na Austrália, muitos tipos de solos apresentam baixos teores de fósforo, elemento essencial para o crescimento dos vegetais. As florestas apresentam diferentes composições (fechadas ou abertas) e taxas de crescimento, resultando em ambientes distintos, os quais sustentam diversas populações de animais como insetos e pássaros [2, 3].

As centenas de espécies de *Eucalyptus* apresentam grande diversidade de propriedades, embora somente algumas estejam compreendidas entre aquelas mais popularmente utilizadas. A seleção da espécie para o plantio é realizada de acordo

com as condições de clima e solo, sendo mais recomendadas quanto maior a semelhança das condições do local de plantio com o local de origem. Como as espécies ocorrem em uma grande faixa latitudinal da Austrália, são didaticamente classificadas em função da região climática que ocupam [3].

A disseminação de sementes de *Eucalyptus* no mundo começou no início do século XIX. Na América do Sul, o primeiro país a introduzir o eucalipto foi o Chile, em 1823 e, posteriormente, a Argentina e o Uruguai. Por volta de 1850, países como Portugal, Espanha e Índia iniciaram o plantio de *Eucalyptus* [4]. No Brasil, conforme melhor detalhado na Sessão 2 desta Tese - Biotecnologia e História da Cultura do *Eucalyptus* no Brasil - árvores de eucalipto foram introduzidas em 1868. As principais espécies de eucaliptos plantadas no Brasil são de origem tropical e subtropical [3]. A seguir, segue uma breve descrição das espécies de eucaliptos utilizadas criticamente em programas de melhoramento genético no Brasil e que foram utilizadas na condução de experimentos constituintes da presente tese de doutorado.

A área de ocorrência natural do *E. grandis*, na Austrália, estende-se em forma descontínua e fragmentada por uma longa faixa costeira, desde Newcastle até Atherton. Embora estas duas localidades estejam situadas em diferentes regiões latitudinais, a primeira temperada e a segunda tropical, não existem diferenças climáticas apreciáveis entre ambas, já que no setor meridional os povoamentos estão situados quase ao nível do mar, e no setor setentrional estão entre 600 e 1.250 m de altitude [4]. O clima de toda esta área varia de temperado-quente a subtropical-moderado, com invernos suaves e chuvas abundantes e bem distribuídas, menos no setor norte, onde são de tipo periódico [3]. Quando plantadas em regiões tropicais úmidas, as árvores de *E. grandis* tornam-se sensíveis ao fungo *Cryphonectria cubensis*, agente causal da doença conhecida como “cancro do eucalipto” [7].

O *E. grandis* é uma espécie que possui qualidades excelentes para a produção de madeira, celulose e papel, superando qualquer outra em incremento, quando as condições ambientais são adequadas, sendo esta a causa de sua grande aceitação. Entre suas várias características destacam-se o hábito de se desramar espontaneamente, o que dá origem a fustes lisos com aspecto colunar. Sua madeira é

boa para serraria e excelente para a produção de celulose. Existem restrições sobre seu poder de rebrota depois de dois cortes periódicos, sendo inferior a *E. saligna* neste aspecto. É a espécie mais plantada no Brasil, especialmente na forma de híbridos com *E. urophylla*. Devido à sua plasticidade genética, árvores de *E. grandis* são muito utilizadas para obtenção de híbridos e na clonagem de árvores selecionadas [3, 8].

E. pellita tem, na Austrália, duas áreas de ocorrência: uma na faixa costeira sul de Quennsland até Nova Gales do Sul, região esta que corresponde ao sul do Brasil; e a segunda área está situada no norte costeiro de Queensland, sendo caracterizada por um clima tropical com precipitações entre 1.250 e 2.000 mm anuais, de tipo periódico e com um período seco de 5 a 7 meses de duração [9]. Sua madeira de cor vermelha escura é ótima para serrarias, pois é mais densa que as espécies mais usuais no Brasil [*E. grandis, E. saligna* e (*E. urophylla* x *E. grandis*)], apresentando valores por volta de 0,55 a 0,6 g/cm^3. Esta característica confere muita similaridade às madeiras de *E. tereticornis* e *E. urophylla*, pois seu teor de lignina é alto, cerca de 28%, e possui menores rendimentos de celulose, com rendimento inferior em cerca de 2% àquele encontrado no híbrido (*E. urophylla* x *E. grandis*)[10].

Nativo de algumas ilhas orientais do arquipélago de Sonda (Timor, Flores, Adonara, Lomblem, Pantar, Alor e Wetar) situadas ao norte da Austrália, o *E. urophylla* tem, em Timor, sua maior distribuição altitudinal, ocorrendo em colinas e montanhas entre 550 e 2.940 m, com clima que varia desde tropical subúmido a montano úmido [11]. O *E. urophylla* é de floresta aberta alta com casca lisa e/ou fibrosa. As árvores têm de 30 a 60 m de altura e a penetração de luz permite a formação de subbosque. A espécie apresenta fuste com boa forma, densidade básica ao redor de 0,5 g/cm^3 e lignotubérculo que é responsável pela sua boa brotação. Responde ao espaçamento e à adubação, apresenta bom desenvolvimento na segunda rotação, possui resistência ao défict hídrico, é suscetível à geada e apresenta variações clinais e ecotípicas importantíssimas para o melhoramento da espécie [12].

O interesse pelo *E. urophylla* surgiu no Brasil nos últimos anos depois de comprovada sua alta resistência ao cancro do eucalipto, com maior potencial de crescimento em termos de área plantada. Esta característica, como também as

propriedades de sua madeira, indica que o *E. urophylla* pode ser um bom substituto do *E. grandis* nas localidades onde este último torna-se suscetível ao cancro. Apresenta boa produtividade de madeira e potencial de utilização para os mais diversos fins como a fabricação de celulose e papel, chapas duras, serraria e produção de carvão [2, 3].

E. dunnii é uma espécie que ocupa uma área restrita no nordeste de Nova Gales do Sul (latitude de 28°S a 30°S), com altitude entre 300 a 800 m, uma região de baixas altitudes e fundo de vale. Pelas características de sua área de origem, é potencialmente apta para o Sul do Brasil. No Planalto Sul, demonstrou discreta resistência ao frio, suportando temperaturas mínimas de até -5°C. A produção de sementes tem sido uma restrição para a ampliação do uso desta espécie [3]. *E. dunnii* respondeu bem ao crescimento em volume e altura quando introduzido nas regiões brasileiras onde ocorrem chuvas de verão bem distribuídas. Por este motivo, esta espécie tem despertado o interesse das empresas de celulose, especialmente pela sua densidade, o que traduz em maior rendimento no processo industrial [13].

Natural das montanhas dos Estados de Victoria e Nova Gales do Sul na Austrália, e da parte leste da Tasmânia, sempre ao longo das costas, o *E. globulus* é uma das árvores mais amplamente cultivadas na Austrália, com possibilidade de êxito no planalto Sul do Brasil. Possui madeira rígida e dura, acinzentada, com tronco ereto e galhos de caule longo e fino. As toras são fortes e duráveis e utilizadas para os mais diversos fins como dormentes para construção de estradas de ferro, produção de papel, óleo e mel [3]. *E. globulus* possui grande valor comercial pelo alto teor de celulose e baixo teor de lignina e tem sido amplamente cultivado em regiões de inverno chuvoso e temperaturas relativamente baixas, ocupando aproximadamente 22% da área total de *Eucalyptus* no mundo, sendo amplamente cultivado em Portugal, Chile, Espanha [4] e Uruguai [14] e, em áreas menores, na Bolívia, China, Colômbia, Etiópia, Peru e EUA [15].

Outra importância comercial de destaque do *E. globulus* é seu uso na indústria farmacêutica. As folhas desta espécie são ricas em tanino e óleo essencial, tendo emprego na fabricação de produtos farmacêuticos inalantes, estimulantes da

secreção nasal, produtos de higiene bucal, ou simplesmente como função de dar sabor e aroma aos medicamentos. Os óleos também estão presentes nas indústrias de perfumes e nas indústrias de produtos de limpeza. Na indústria de pasta de celulose para a produção de papel de fibra curta, é conhecida a importância crescente de *E. glubulus* para a fabricação de uma vasta gama de papéis, como papéis para impressão, escrita e domésticos [16].

E. saligna é uma espécie muito próxima ao *E. grandis* nos aspectos botânicos, ecológicos e silviculturais. Sua área natural encontra-se na região litorânea e nos vales das cadeias montanhosas próximas ao litoral de New South Wales e ao sul de Queensland [17, 18]. Sua distribuição ocorre em forma descontínua de 21° a 36° de latitude Sul, desde o nível do mar até 1.000 m. A precipitação pluviométrica ideal para o cultivo de *E. saligna* situa-se entre 800 e 1.200 mm, com chuvas distribuídas durante o ano ou concentradas no verão, porém a estação de seca não deve ultrapassar quatro meses [18]. Com relação às temperaturas, *E. saligna* apresenta alta suscetibilidade a geadas. Porém, esta espécie adapta-se facilmente em locais onde a temperatura média nas máximas do mês mais quente encontra-se na faixa de 28 a 30°C e das mínimas do mês mais frio entre 3 a 4°C [19]. No Brasil, seu plantio é recomendado para todas as regiões, com restrições a locais onde ocorram geadas ou deficiências hídricas severas [18].

E. saligna fornece madeira clara de baixa densidade, apta para a produção de celulose e numerosas outras finalidades. Sua madeira é indicada para usos gerais como em construções, laminações de móveis, caixotaria, postes, escoras, mourões, celulose e carvão, e para reflorestamento [3].

5. A Biotecnologia e a História da Cultura de *Eucalyptus* no Brasil

Historicamente, as primeiras mudas de eucalipto (*Eucalyptus* sp.) introduzidas no Brasil foram plantadas no Rio Grande do Sul em 1868 e, no mesmo ano, alguns exemplares foram plantados na Quinta da Boa Vista, Rio de Janeiro [20]. Entretanto, os primeiros estudos científicos com *Eucalyptus* no Brasil foram realizados por Edmundo Navarro de Andrade na Antiga Companhia Paulista de Estradas de Ferro em 1904, cujo objetivo principal foi atender à demanda de madeira para a construção de ferrovias. Estes trabalhos, considerados avançados para a época, selecionaram várias espécies promissoras. Entretanto, as plantações eram muito heterogêneas, resultando em baixas produtividades. Assim, em 1941, Carlos Arnaldo Krug do Instituto Agronômico de Campinas (IAC) foi convidado para iniciar um programa de melhoramento genético das diversas espécies de *Eucalyptus* disponíveis [21].

Nos primeiros anos da década de 1940, a Cia. Melhoramentos iniciou a condução de pesquisas com o objetivo de obter celulose a partir de eucalipto. Os estudos, promovidos por Hasso Weiszflog, obtiveram sucesso em setembro de 1946, resultando na produção de 45.400 kg de celulose até o final daquele ano. A partir do mesmo ano, a empresa centrou esforços para a produção de celulose branqueada e papel para escrever, a partir de fibras de eucalipto [3].

Em julho de 1947, a Cia. Melhoramentos obteve novo sucesso, produzindo 3.400 kg de celulose de alto padrão. Foram realizadas inúmeras experiências, com proporções variáveis de celulose de eucalipto, visando a substituição da pasta e celulose de álamo (*Populus tremuloides*) que, até então, era utilizada para a fabricação de papel. Em 30 dias, a empresa produziu cerca de 30.000 kg de celulose de eucalipto para a produção de papel comercial [2, 3].

Com o objetivo de revender papéis nacionais e importados, foi fundada, em 1924, a Leon Feffer e Cia. A produção de papel própria começou em 1942, com a instalação da primeira máquina de papel, que utilizava como matéria-prima celulose

importada. Em 1951, a empresa iniciou pesquisas que resultaram na obtenção da celulose de eucalipto, uma alternativa nacional à importação [3].

O nome da Cia. Suzano surgiu em 1955, quando foi incorporada ao grupo da indústria Papel Euclides Damiani. Dois anos depois, graças aos esforços de uma equipe liderada por Max Feffer, a empresa firmava seu pioneirismo no mercado internacional. Em agosto de 1957, a Cia. Suzano iniciou a produção industrial de papel com 100% de celulose de eucalipto, após uma série de ensaios de cozimento, lavagem e branqueamento. Estes ensaios foram realizados por profissionais da própria empresa, e na Universidade da Flórida (Gainesville, EUA), que comprovou os resultados obtidos no Brasil [2, 3].

Até 1966, foram plantadas 470.000 hectares (ha) de eucaliptos em todo o Brasil, sendo que 80% dessa área localizava-se no Estado de São Paulo. Naquele ano, o governo brasileiro criou o Programa de Incentivos Fiscais, intensificando o plantio de florestas em várias regiões do país. Segundo informações do então Instituto Brasileiro de Desenvolvimento Florestal (IBDF), os projetos com incentivos fiscais totalizaram, até dezembro de 1987, ano de término dos incentivos fiscais, cerca de 6 milhões de ha, sendo 52% com eucalipto [2].

Em 1964, o Instituto Florestal do Estado de São Paulo assumiu as atividades desenvolvidas pela Antiga Companhia Paulista de Estradas de Ferro. Nesta época havia um programa intensivo de melhoramento genético, por meio da seleção de árvores superiores e, além disto, criou-se um programa de certificação de sementes. Preocupadas em atender à demanda por sementes melhoradas e conhecer melhor as espécies/procedências de eucalipto, algumas empresas florestais privadas criaram o Instituto de Pesquisas e Estudos Florestais (IPEF), sediado na Escola Superior de Agricultura Luís de Queiroz da Universidade de São Paulo (*campus* de Piracicaba), em 1968. Todavia, as sementes produzidas e comercializadas possuíam baixa qualidade genética e fisiológica [21].

No final da década de 1960, a cultura do *Eucalyptus* expandiu-se para outras regiões. Nesta década surgiu o primeiro grande problema da cultura, como já referido, o cancro causado pelo fungo *C. cubensis* Burner (Hodges). Nesta época,

também se descobriu o grande potencial dos híbridos (*E. grandis* x *E. urophylla*), que associavam resistência ao cancro e características silviculturais vantajosas [21].

Diante desse cenário, em 1969, as empresas Champion e Duratex trouxeram para o Brasil o Professor Lindsay Dixo Pryor, da Austrália. Este pesquisador constatou que, devido principalmente à falta de isolamento geográfico em que foram submetidas às introduções de Navarro de Andrade, houve intensa hibridização interespecífica, levando à descaracterização botânica e à perda de vigor na segunda geração. Assim, introduziram-se novas procedências da Austrália, do Zimbabwe e da África do Sul, visando formar novas populações para os programas de melhoramento genético [2, 3, 21].

Face ao sucesso e ao pioneirismo desse tipo de iniciativa, ações semelhantes foram desenvolvidas em universidades com o apoio de empresas do setor privado, com a criação da Sociedade de Investigações Florestais (SIF) junto à Universidade Federal de Viçosa, em Viçosa, MG, e da Fundação de Pesquisas Florestais (FUPEF), associada à Universidade Federal do Paraná, em Curitiba, PR. Neste mesmo período foi criado o Centro Nacional de Pesquisa Florestal (CNPF) da Empresa Brasileira de Pesquisa Agropecuária (Embrapa), hoje Embrapa Florestas, em Colombo, PR. Estes institutos lideraram a instalação de uma rede de ensaios nas áreas de Melhoramento Genético e Silvicultura, os quais serviram de base para que o Setor Florestal Brasileiro obtivesse aumentos significativos na produtividade dos maciços florestais de eucaliptos, como o ilustrado na Figura 1. Em 1975, as empresas Champion e Duratex, em trabalho conjunto com o IPEF, selecionaram áreas para a produção de sementes melhoradas geneticamente. Aliadas a este fato, as pesquisas demonstraram, por exemplo, que o eucalipto respondia positivamente à adubação [3].

Figura 1 – Vista aérea de plantio comercial de *Eucalyptus*.
Fonte: http://www.portaldoagronegocio.com.br/conteudo.php?id=45098

A década de 1980 foi marcada inicialmente pelos trabalhos da empresa Aracruz, que dominou a técnica de propagação vegetativa do eucalipto. Plantios clonais foram efetuados e, novamente, conciliando o melhoramento genético (utilização de híbridos) e adubações diferenciadas, a Aracruz produziu florestas com produtividade de 50 m^3/ha/ano [3]. Houve grande fortalecimento da área de pesquisa e desenvolvimento, proporcionado pelos investimentos realizados pelas empresas privadas que criaram seus próprios departamentos técnicos. As universidades e institutos de pesquisas continuaram sua missão de liderança na condução de pesquisa básica.

O período de 1980 a 1995 caracterizou-se pela redução dos investimentos em estratégias de melhoramento genético, dando-se grande ênfase à propagação clonal. O híbrido interespecífico (*E. grandis* x *E. urophylla*), popularmente denominado de "*E. urograndis*", passou a receber atenção especial. Neste período, em que todos os esforços foram dedicados à estaquia, as empresas buscaram adaptar seus genótipos às novas regiões geográficas [22].

Com a expansão da área reflorestada, principalmente no nordeste brasileiro, a caracterização do solo, o seu melhor manejo e a utilização de espécies/procedências de eucaliptos adequadas aos locais específicos passaram a ter

fundamental importância nas recomendações técnicas para o plantio em escala comercial. Nos anos 1990, as atenções voltaram-se para a utilização racional dos recursos naturais, procurando-se preservar, conservar e interligar as áreas naturais, manter a produtividade florestal, promover o uso múltiplo das florestas e desenvolver sistemas e equipamentos visando minimizar os esforços físicos dos trabalhadores, riscos com acidentes e possíveis danos ao solo. Neste período, as empresas florestais adotaram novos modelos de gestão, passando por um processo de reengenharia e terceirização de atividades. A produção de madeira de eucalipto para outras finalidades, principalmente serraria, passou a ser considerada com maior ênfase pelas empresas florestais. O uso múltiplo das plantações de eucalipto começou a despertar interesse e viabilidade. As empresas Duratex, Caf e Klabin investiram para dominar a tecnologia de processamento de eucalipto em serraria de modo a valorizar a madeira [2]. Em 1999, a Aracruz iniciou operações de uma serraria de avançada tecnologia localizada no sul da Bahia, com capacidade de produção de 800.000 m^3/ano, projetada para fabricar, a partir de plantações florestais de eucalipto, produtos sólidos de madeira destinados a indústria de móveis e de construção civil, do Brasil e do exterior [2].

Nas últimas três décadas, houve a implementação e parcial substituição da macropropagação pela micropropagação, devido aos inconvenientes da primeira. As empresas perceberam também o equívoco do abandono dos programas de melhoramento genético e intensificaram os programas de seleção com grande ênfase à seleção recorrente recíproca, visando melhorar suas populações, para a obtenção de híbridos interespecíficos superiores [21].

Um novo desafio para o melhoramento do eucalipto está na integração da biotecnologia mais avançada ao seu cultivo, o que compreende a identificação de genes controladores das características de importância econômica e ambiental e a transferência destes genes entre árvores por meio de cruzamentos controlados ou modificação direcionada. O uso da biotecnologia no melhoramento genético florestal envolveu, até 2000, basicamente os métodos de cultura e propagação de plantas e tecidos *in vitro* e o emprego de marcadores moleculares.

As técnicas de cultura de tecidos são utilizadas visando solucionar problemas em diferentes segmentos da área agrícola e vegetal [23]. Em programas de melhoramento florestal, a propagação vegetativa *in vitro*, também denominada de micropropagação, tem especial destaque porque permite a manutenção e multiplicação rápida de mudas, a partir de um genótipo superior, em períodos de tempo e espaço físico reduzidos, além de assegurar um material livre de patógenos.

A geração de árvores geneticamente modificadas (GM) por procedimentos de transgenia representa o próximo passo dessa evolução, na medida em que permite a criação de plantas de crescimento mais rápido, ambientalmente mais econômicas, e que dão origem a madeiras com características ainda mais adequadas ao beneficiamento industrial e menos impactantes ao meio ambiente. Esta tecnologia possibilita também melhorias introduzidas nas fibras, nas composições químicas e nas propriedades físicas da madeira, favorecendo o crescimento volumétrico, a tolerância a estresses ambientais como frio, salinidade e estresse hídrico, a resistência a pragas e a doenças, e o uso racional dos recursos naturais. Ao contrário do que ocorre nas culturas agrícolas, nas quais os genes provêm de outros organismos, as espécies de eucaliptos GM têm sido desenvolvidas com genes encontrados no próprio gênero *Eucalyptus*, dada sua ampla variabilidade [24].

A transformação genética abriu uma nova perspectiva aos programas de melhoramento, ampliando e disponibilizando novos genes para determinadas características impostas pela incompatibilidade sexual ou pela variabilidade genética [18]. Neste contexto, a transformação genética, aliada à técnica de cultura de tecidos vegetais, obteve avanços consideráveis nos últimos anos. De acordo com a Organização das Nações Unidas para Agricultura e Alimentação (FAO), mais de 210 ensaios de campo estão sendo conduzidos em 16 países, envolvendo aproximadamente 15 espécies [25]. Destas, o gênero *Populus* é o mais estudado com 51% dos ensaios, seguido pelo *Pinus*, com 23%, pelo *Liquidambar*, com 11%, e pelo *Eucalyptus*, com 7%. A maioria dos ensaios com árvores GM, em torno de 64%, está sendo avaliada nos EUA, seguido pela França e Finlândia, na Europa. No Brasil, mais de 15 liberações planejadas no meio ambiente de árvores GM de eucalipto foram autorizadas pela Comissão Técnica Nacional de Biossegurança (CTNBio) [24].

Das principais modificações genéticas realizadas em árvores, a metade está relacionada a testes dos métodos de transformação genética e a outras questões de biologia básica. Dos ensaios remanescentes, 13% foram dedicados à avaliação da tolerância a herbicidas, 12% à resistência biótica, 9% à química da madeira e 6% ao estudo de fertilidade. Hoje, as características principais em foco são: aumento da densidade da madeira, redução e modificação no conteúdo e no tipo de lignina e aumento no conteúdo de celulose [24].

A tecnologia de marcadores moleculares, aliada às técnicas clássicas do melhoramento de eucalipto, contribuiu significativamente para o conhecimento básico da cultura, das características estudadas e para a geração e o desenvolvimento de produtos melhorados desde meados da década de 1990. Assim, os marcadores moleculares foram e permanecem ferramentas auxiliares no melhoramento genético florestal permitindo, entre outros, avaliações da adequação da base genética de populações florestais para a utilização em programas de melhoramento. Praticamente todas as técnicas-chave de marcadores moleculares foram empregadas no melhoramento genético de *Eucalyptus*, destacando-se os polimorfismos de tamanhos de fragmentos de DNA gerados por restrição (RFLP, do inglês, *Restriction Fragment Length Polymorphism*), DNA polimórfico amplificado aleatoriamente (RAPD, do inglês, *Random Amplified Polymorphic DNA*), polimorfismos de tamanho de fragmentos amplificados (AFLP, do inglês, *Amplified Fragment Length Polymorphism*), microssatélites ou sequências simples repetitivas (SSR, do inglês, *Single Sequence Repeat*) e polimorfismos de um único nucleotídeo (SNP, do ingês, *Single Nucleotide Polymorphism*). Tais técnicas foram revezando-se em virtude dos avanços tecnológicos e facilidades operacionais a ponto de, hoje, fazer-se uso de procedimentos de alto desempenho e capazes de gerar volumes imensos de marcadores moleculares como, por exemplo, *deep sequencing* de sequências expressas ou hibridizações de microarranjos de DNA [26].

Além da introdução de plantas transgênicas e marcadores moleculares como ferramentas modernas de melhoramento genético, centros de pesquisa e empresas investiram fortemente em projetos genoma de eucalipto, a partir de 2000. No Brasil, dois projetos de pesquisa do transcriptoma do *Eucalyptus* foram iniciados

paralelamente. O Projeto ForESTs [27], financiado pela Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) e por quatro empresas de celulose e papel foi concluído em 2002, com a geração de um banco de 120.000 sequências de genes expressos (ESTs, do inglês, *Expressed Sequece Tags*) de *E. grandis*. O Projeto *Genolyptus*, iniciado em 2002 e concluído em 2009, foi concebido por uma parceria entre o governo federal, por intermédio do Ministério de Ciência e Tecnologia, o setor acadêmico e de pesquisa representado por sete universidades e três centros da Embrapa e o setor privado, com 14 empresas florestais, 13 brasileiras e uma portuguesa. O Projeto Genolyptus permitiu a geração de um banco de aproximadamente 100.000 ESTs em várias espécies de *Eucalyptus*, além do mapeamento de milhares de microssatélites, caracterização do padrão de expressão de 21.442 genes por microarranjos de DNA, progênies de diversos cruzamentos, dados de análises da qualidade da madeira, mapeamento genômico, entre outros resultados [26].

Em nível mundial, após o congresso organizado pela *International Union of Forest Research Organizations* (IUFRO) em Hobart/Australia, diversos pesquisadores do assunto uniram esforços para a formação de um consórcio de troca de experiências e conhecimentos no campo da pesquisa avançada em genômica de *Eucalyptus*, denominado EUCAGEN. O consórcio abrigou cientistas do Brasil, Austrália, Japão, Estados Unidos, Portugal, Espanha, França, África do Sul e Bélgica. Como principal fruto deste consórcio, foi recentemente disponibilizado publicamente os genomas de *E. grandis* e *E. globulus*, gerados pelo *Joint Genome Institute* (JGI) (http://www.phytozome.net/eucalyptus.php) do Departamento de Energia (DoE) dos EUA, a partir de espécimes fornecidos pelo Brasil (*E. grandis*) e pela Austrália (*E. globulus*) [28].

6. A Produção de *Eucalyptus* no Brasil

A área de florestas com eucaliptos está em franca expansão na maioria dos Estados brasileiros, com um crescimento de 4,4% da área plantada de um total de 6.310.450 ha de florestas de *Eucalyptus* e *Pinus* plantadas no ano de 2009 (Figura 2), totalizando um acréscimo médio de 152.700 ha ao ano. A expansão na área plantada de eucalipto é resultado de um conjunto de fatores que vêm favorecendo o plantio em larga escala deste gênero, como o rápido crescimento em ciclo de curta rotação, a alta produtividade florestal e a expansão e direcionamento de novos investimentos por parte de empresas de segmentos que utilizam sua madeira como matéria prima em processos industriais; além das previsões de expansão no segmento de celulose e papel [29].

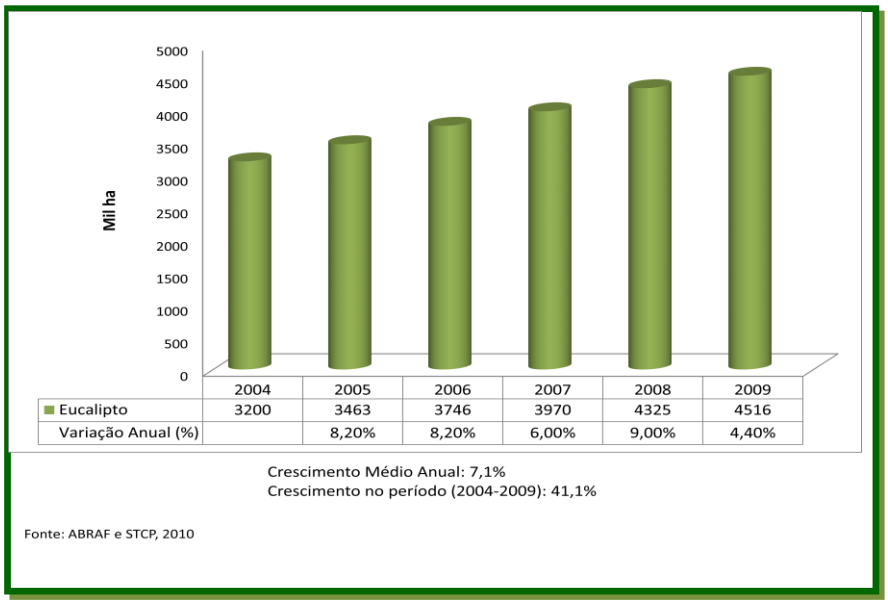


Figura 2 – Evolução da área de florestas plantadas com eucaliptos, no período de 2004 a 2009, no Brasil. Reproduzido de ABRAF e STCP (2010) [29].

Os dados de crescimento das florestas de eucaliptos por Estado da federação no período de 2005 a 2009 mostraram que o maior crescimento ocorreu no

Estado da Bahia, seguido por São Paulo, Maranhão, Mato Grosso do Sul e Minas Gerais, os quais, em conjunto, totalizaram 143.000 ha de incremento das áreas (Tabela 1). Com base nesses dados, constata-se que 56% das áreas com florestas plantadas de eucaliptos no Brasil, até 2009, localizam-se na região Sudeste, com destaque para o Estado de Minas Gerais, com 1.300.000 ha, e São Paulo, com aproximadamente 1.030.000 ha, com participação de 29% e 23% do total do País. Outro Estado que merece destaque é a Bahia, com aproximadamente 628.000 ha de área plantada, representado 14% do total do País. Em conjunto, estes três Estados detêm dois-terços do total plantado com este grupo de espécies no Brasil. O Estado com maior aumento percentual de florestas plantadas em áreas próprias em 2009 foi, entretanto, o Rio Grande do Sul, com crescimento de 42% em relação a 2008 [29].

Tabela 1 – **Área de florestas plantadas com *Eucalyptus* no Brasil (2005-2009), destacando o Estado do Rio Grande do Sul.** Reproduzido de ABRAF e STCP (2010) [29].

UF	Eucalipto (ha)				
	2005	**2006**	**2007**	**2008**	**2009**
Minas Gerais	1.119.259	1.181.429	1.218.212	1.278.210	1.300.000
São Paulo	798.522	915.841	911.908	1.001.080	1.029.670
Bahia	527.386	540.172	550.127	587.610	628.440
Mato Grosso do Sul	113.432	119319	207.687	265.520	290.890
Rio Grande do Sul	179.690	184.245	222.245	277.320	271.980
Espírito Santo	204.035	207.800	208.819	210.410	204.570
Paraná	114.996	121.908	123.070	142.430	157.920
Pará	106.033	115.806	126.286	136.290	139.720
Maranhão	60.745	93.285	106.802	111.120	137.360
Santa Catarina	61.166	70.341	74.008	77.440	100.140
Goiás	47.542	49.637	51.279	56.880	57.940
Amapá	60.087	58.473	58.874	63.310	62.880
Mato Grosso	42.217	46.146	57.151	58.580	61.530
Tocantins	2.124	13.901	21.655	31.920	28.380
Outros	25.285	27.491	31.588	27.580	28.380
TOTAL	**3.462.719**	**3.745.794**	**3.969.711**	**4.325.430**	**4.515.730**

As exportações brasileiras de produtos de florestas plantadas atingiram US$ 5,6 bilhões em 2009, diante de US$ 6,8 bilhões obtidos em 2008, diminuindo, portanto, em 18%. Por outro lado, o setor de florestas plantadas foi responsável por 4% do total de exportações totais do país em 2009, representando um ponto percentual acima da participação em 2008 [29].

A participação brasileira no mercado mundial de produtos florestais foi de 2% no ano de 2008. No mercado de papel, o Brasil foi 12º produtor mundial, com uma produção de 12.700.000 t em 2008, correspondendo a 2,3% da produção global. Por outro lado, no mercado de celulose, o Brasil produziu no mesmo ano 9.400.000 t, o que correspondeu a 6,5% da produção mundial, ocupando a quarta posição [30].

Entre os anos de 2000 e 2008, ocorreu um crescimento médio anual de 20,9% no valor e na quantidade exportada do segmento de celulose e papel do país. Na quantidade produzida, o aumento foi de 3,8% no mesmo período [30]. Considerando um período mais recente, janeiro a abril de 2010, observou-se que as exportações brasileiras de celulose totalizaram US$ 1,13 bilhão no primeiro trimestre, um acréscimo de 48,8% em relação ao mesmo período de 2009. Somados os embarques de papel, a indústria nacional exportou 41,4% a mais no trimestre, com US$ 1,62 bilhão [31].

Os EUA, a Holanda, a China, a Bélgica e a Itália foram os maiores importadores de celulose brasileira, representando, juntos, 80% do total exportado. Com relação ao papel, o maior mercado para o produto brasileiro foi a Argentina, os EUA, o Chile, a Venezuela e o Paraguai, que totalizaram 47% das exportações neste setor [29, 30].

No mercado mundial de produtos de madeira sólida, o Brasil ocupou em 2006 a oitava posição, com participação de 4,1% do total produzido. Das toras de *Eucalyptus* produzidas no Brasil em 2008, 91,12% foram consumidas para a produção de celulose e papel (43,85%), carvão (20,42%) e para lenha (26,85%). O beneficiamento de toras para a obtenção de produtos de maior valor agregado como portas, batentes, pisos, moldura, ferramentas e painéis de sarrafo (*Edge Glued Panel* - EGP) constituiu-se no quarto maior mercado da madeira de *Eucalyptus*, com 3,71% do

total [32]. Com relação ao carvão produzido, cerca de 51,1% foram de origem de floresta plantada, sendo a quase totalidade obtido a partir de *Eucalyptus* [29, 32].

OBJETIVOS

Diante da importância sócio-econômica que a silvicultura desempenha no mercado brasileiro e do aumento progressivo das áreas plantadas com florestas de *Eucalyptus*, conforme acima exposto, este estudo foi desenvolvido com o objetivo geral de contribuir cientificamente à biotecnologia dedicada ao melhoramento genético de *Eucalyptus grandis*, visando permitir avanços científico e tecnológico ainda mais amplos, que comprovadamente, têm propiciado incrementos significativos da produtividade de madeira e derivados.

Três objetivos específicos foram propostos para a composição do presente estudo. Em primeiro lugar, realizou-se o levantamento de genes expressos em flores abertas e botões florais em antese de *E. grandis*, por intermédio da construção de bibliotecas de expressão de flores e o sequenciamento automático de cDNAs. Para tanto, os botões florais foram divididos em quatro partes: sépalas/pétalas; anteras/estames; pistilos e carpelos/receptáculos florais. Como forma de validação da representatividade de genes expressos em cada parte, algumas das sequências obtidas fora avaliadas quanto ao perfil transcricional pela técnica de reação em cadeia da DNA polimerase em tempo real ou quantitativa precedida de transcrição reversa (RT-qPCR).

Dentre os genes identificados em flores de *E. grandis*, algumas famílias de fatores de transcrição destacaram-se. Foram encontrados dois genes pertencentes à família de fatores de transcrição de domínio único de ligação ao DNA (Dof, do inglês, *DNA one finger*). Assim, um segundo objetivo específico, que se tornou alvo de um estudo mais aprofundado, foi a caracterização de genes de *E. grandis* potencialmente codificadores de fatores Dof. Estes fatores, exclusivos do reino vegetal, são conhecidos por atuarem como ativadores ou repressores transcricionais em diversos metabolismos das plantas e a respostas a estímulos do

ambiente. Até o presente momento, poucos relatos existem na literatura sobre estes fatores de transcrição em plantas arbóreas.

Durante a execução deste trabalho, notou-se a dificuldade de se utilizar genes normalizadores já descritos na literatura para outras espécies de plantas para experimentos de avaliação da expressão gênica no gênero *Eucalyptus*. Diante deste quadro, o terceiro objetivo específico que levou ao desenvolvimento do presente trabalho, foi selecionar genes por métodos estatísticos confiáveis, validar o perfil constitutivo destes genes por RT-qPCR e analisar a robustez dos dados obtidos. Estas análises tornaram-se fundamentais para a obtenção de cálculos mais confiáveis nas análises de expressão gênica em tempo real.

CAPÍTULO I

Estudo do Transcriptoma em Flores de *Eucalyptus grandis*

Fundamentação Teórica

Eucalyptus grandis e Anatomia da Flor

As inflorescências de *E. grandis* são em umbelas axilares, com pedúnculo achatado, contendo de 6 a 12 flores brancas. Os botões florais são sésseis, piriformes, com opérculo ligeiramente apiculado. Os frutos (cápsulas) são piriformes, em geral verde-azulados, deiscentes, com valvas encurvadas apresentando 7 mm de diâmetro, com sementes marrons pequenas. Esta espécie é muito semelhante ao *E. saligna*, distinguindo-se pelos botões verde-azulados e pelas valvas encurvadas das cápsulas [8, 33].

A organização básica da flor das angiospermas é extremamente conservada entre todas as espécies. Os órgãos centrais de reprodução (estames e carpelos) são sempre cercados por estruturas do perianto (sépalas e pétalas). Alterações no número, forma tamanho, cor e disposição destes órgãos florais, geralmente devido a adaptações estratégicas particulares à polinização levaram à evolução de uma enorme variedade de estruturas florais [34].

E. grandis apresenta flores hermafroditas, pequenas (0,5 cm), de coloração branca e em número maior que sete em inflorescência do tipo umbela [35, 36] tendo os insetos como agentes polinizadores. O surgimento dos botões caracteriza o início da reprodução, fase esta que se dá a partir dos 2 a 3 anos de idade [4, 36]. O número de botões florais é, em média, superior a 20.000 por indivíduo [37]. *Eucalyptus* são naturalmente protândricos (pólen disponível antes do estigma estar receptivo), apresentando um sistem misto de reprodução em que prevalece a feculdação cruzada, portanto, provavelmente esta seja adaptação para promover a polinização cruzada [4, 38]. A protandria minimiza a autofecundação, porém não a inibe totalmente, pois as plantas apresentam flores ao logo do ano e resultam em quatro classes de sementes:

(i) autofecundadas – em média 30% [39, 40], que resultam em maior mortalidade das plantas, crescimento lento e forma ruim do fuste [4]; (ii) heterozigotos em poucos locos pelo cruzamento com árvores da mesma família (meias irmãs); (iii) hterozigotos em vários locos resultantes do cruzamento com árvores não aparentadas; (iv) híbridos interespecíficos resultantes do cruzamento com espécies afins como *E. saligna*, *E. pellita*, *E. resinifera*, *E. urophylla*, *E. tereticornis* e *E. camaldulensis* [36].

Os botões florais de *Eucalyptus* são protegidos pelo opérculo, derivado da fusão das pétalas e sépalas (Figura 1). O desprendimento do opérculo na flor madura é seguido imediatamente pelo espalhamento dos estames e antese, enquanto o estigma não se torna receptivo até vários dias depois. O tempo entre a remoção do opérculo (e a antese) até o início da receptividade em *E. grandis* é de 4 a 6 dias após a antese [40] A sequência da antese e a receptividade também são afetadas por condições ambientais como, por exemplo, as temperaturas altas, quando o processo de antese ocorre mais rapidamente [40]. Flores individuais de uma inflorescência e dentro de uma mesma árvore variam em seu tempo de florescimento para que haja uma ampla oportunidade para autopolinização, particularmente para as flores abertas tardiamente dentro da coroa de flores *de Eucalyptus* [4].

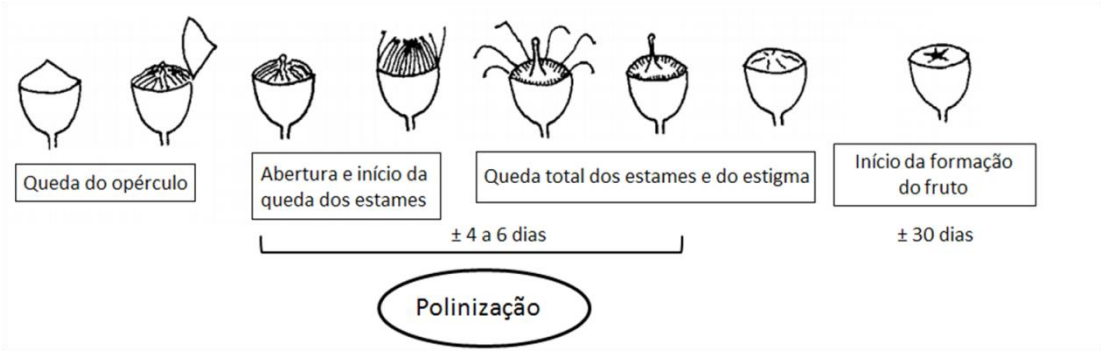


Figura 1 – Esquema dos estágios do florescimento e frutificação de *E. grandis*. Adaptado de Souza e Higa [10].

A primeira descrição botânica do gênero sob o nome de *Eucalyptus* foi feita em 1788 pelo botânico francês Charles-Louis L´Héritier de Brutelle [41]. O nome genérico *Eucalyptus* é derivado do "*eu*", que significa "boa", e "*calyptus*", que significa "cobertura", referindo-se exatamente ao opérculo que cobre o estigma e os estames até que o mesmo caia e as flores se abram. A abertura da flor de *Eucalyptus* é um evento muito particular, denominado antese, caracterizado pela queda do opérculo, composto pelas sépalas e pétalas modificadas em uma estrutura cônica; e a descompactação dos estames (Figura 2). Somente a partir deste momento é que se considera a flor totalmente aberta. Portanto, os típicos eventos moleculares descritos usualmente para pétalas ou tépalas em espécies de plantas ornamentais como cravo (*Dianthus caryophyllus*) [42], lírio (*Hemerocallis hybrid*) [43] e rosa (*Rosa hybrid*) [44] determinam a queda dos estames e do estigma após a fertilização da flor em *E. grandis*.

Figura 2 – Exemplo ilustrativo do evento de antese em botão floral de *E. robusta*
(foto: http://git-forestry-blog.blogspot.com/2008/02/discovering-eucalyptus-anthesis-i.html)

O Transcriptoma de *Eucalyptus*

A transcriptômica, parte das práticas que caracterizam a genômica e a bioquímica funcional , é o estudo do conjunto de transcritos expressos em uma célula, tecido, órgão ou parte de um organismo, em um dado momento fisiológico ou em resposta a algum sinal. Ela representa uma estratégia de caracterização funcional em massa dos genes e de grupos gênicos funcionalmente associados. A análise detalhada do transcriptoma pode tornar possível a identificação e a compreensão de complexas redes metabólicas a que estão sujeitos processos fisiológicos, desenvolvimento e resposta a estímulos [45].

Após a identificação e a anotação dos genes, a prioridade usual passa a ser o estabelecimento dos padrões de expressão para fins de confirmação de suas funções. Apesar de serem bem definidos, sensíveis e robustos, os métodos tradicionais de análise de expressão gênica como *Northern blot*, hibridização *in situ*, reação em cadeia da DNA polimerase (PCR, do inglês, *Polymerase Chain Reaction*) ou PCR quantitativa (em tempo real) precedida de transcrição reversa (RT-qPCR, do inglês, *Reverse Transcription - Quantitative, real-time PCR*), são adequados para a análise de um número pequeno de genes. O estabelecimento de perfis transcricionais proporciona um enorme avanço, pois permite uma análise eficiente de um grande número de genes ao mesmo tempo [46].

As metodologias utilizadas para estabelecer perfis transcricionais baseiam-se na geração de populações de DNA complementar (cDNA) a partir da uma população de RNA mensageiro (mRNA) presente. Estas metodologias podem ser divididas em métodos globais (abertos) e métodos limitados (fechados) de análise. Os métodos globais permitem acesso a potencialmente todos os transcritos presentes em um determinado momento porque não requerem informações prévias de sequências gênicas para a espécie. Já os métodos fechados são limitados pela quantidade de informação disponível para a espécie, uma vez que utilizam sondas e/ou *primers* específicos [45].

Existem diferentes metodologias que permitem a análise global do transcriptoma. Entre elas, as mais utilizadas são o sequenciamento automático e a geração de marcas ou etiquetas de sequências expressas (ESTs, do inglês, *Expressed Sequence Tags*) [47], a amostragem diferencial (DD, do inglês, *Differential Display*) [48], a análise serial da expressão gênica (SAGE, do inglês, *Serial Analysis of Gene Expression*) [49], os polimorfismos de tamanhos de fragmentos amplificados a partir de cDNAs (cDNA-AFLP, do inglês, *cDNA-Amplified Fragment Length Polymorfism*) [50, 51], a hibridização subtrativa por supressão (SSH, do inglês, *Supression Subtractive Hybridization*) [52] e o sequenciamento de assinaturas paralelas em massa (MPSS, do inglês, *Massively Parallel Signature Sequencing*) [53].

O sequenciamento automático de cDNAs ou mRNAs gerando as chamadas ESTs é o mais profuso e informativo método de identificação de genes e de avaliação da expressão gênica em massa. Para a identificação de genes, ele justifica-se pela existência de grandes regiões não codificadoras no DNA, permitindo sequenciar apenas as regiões codificadoras. A base do método para fins de avaliação da expressão gênica consiste no fato de que o nível de mRNA tecido-específico presente em uma espécie é refletido pela frequência da ocorrência de suas ESTs correspondentes em uma biblioteca de cDNA [47].

A partir do conceito original de ESTs, para ajudar na identificação de genes diferencialmente expressos em um tecido ou condição específica, inicia-se uma nova era no campo da genética de árvores, seguindo o exemplo de outras plantas como *Arabidopsis*, milho e arroz. A partir de 1998, iniciado por dois estudos pioneiros descrevendo 5.692 ESTs de *Populus* [54] e 1.000 ESTs de *Pinus* [55], o número de ESTs de árvores tem aumentado progressiva e geometricamente, sendo que atualmente já se encontram disponíveis nos bancos de dados públicos (http://www.ncbi.nlm.nih.gov/dbEST). Como exemplo, podemos citar mais de 500.000 ESTs de *Citrus sp.*, 450.000 de *Pinus sp.*, 430.000 de *Populus sp.*, e 11.000 de *Malus* sp. Para espécies arbóreas, os bancos de dados disponibilizam 213.830 ESTs de *Citrus sinensis*, 328.662 de *Pinus taeda*, 92.142 de *Populus trichocarpa* e 4.955 de *Malus x domestica*. No caso de *Eucalyptus sp.*, o número de ESTs depositadas é muito pequeno

quando comparado com outras arbóreas de importância econômica, com aproximadamente 37.000 ESTs. *E. dunnii* é a espécie de *Eucalyptus* com o maior número, 19.841 ESTs, posi sua madeira é muito semelhante a do E. grandis, podendo ter as mesmas utilizações.

Visando ampliar o número de ESTs disponíveis para o gênero *Eucalyptus*, foram conduzidos no Brasil, a partir de iniciativas de diversas empresas do setor florestal juntamente com universidades e órgãos estaduais ou nacionais de fomento à pesquisa, dois projetos de sequenciamento de ESTs de diferentes espécies e tecidos de *Eucalyptus*. O Projeto *ForESTs*, financiado pela Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) e quatro empresas do setor de celulose e papel, permitiu a geração de 123.889 ESTs [27] e o Projeto *Genolyptus*, financiado pelo Ministério da Ciência e Tecnologia e 14 empresas de celulose e papel, resultou no sequenciamento de aproximadamente 100.000 ESTs [26].

O Projeto *Genolyptus*: Rede Brasileira de Pesquisa do Genoma de *Eucalyptus*, foi oficialmente lançado em fevereiro de 2002 pelo Ministério da Ciência e Tecnologia, que contou com a participação de 10 centros de pesquisas e 14 empresas. O objetivo principal consistiu no sequenciamento, mapeamento e determinação da função de genes de importância econômica para as espécies de *Eucalyptus*, visando a incorporação de tecnologias de genômica em programas de melhoramento e produção florestal em busca de plantas de *Eucalyptus* mais produtivas, mais resistentes a pragas e mais tolerantes a estresses abióticos.

O sequenciamento do transcriptoma de *Eucalyptus* realizado por este projeto, foi feito a partir de 13 bibliotecas de expressão, sendo elas: xilemas de (i) *E. grandis*, (ii) *E. globulus*, (iii) *E. urophylla* e (iv) *E. pellita*; (v) uma mistura de xilemas de nove espécies e híbridos; (vi) uma mistura de floemas de nove espécies e híbridos; (vii) folhas maduras de *E. grandis*; (viii) folhas jovens de *E. grandis*; (ix) folhas jovens de *E. grandis* infectadas a campo pelo fungo *Puccinia psidii*; (x) plântulas *in vitro* de *E. grandis* não tratadas; (xi) uma mistura de plântulas de *E. grandis* submetidas a 20 diferentes tratamentos; (xii) raízes de diferentes espécies e híbridos de *Eucalyptus*; e (xiii) flores e botões florais inteiros de *E. grandis* (Figura 3). Um total de 96.493 ESTs

válidas foi depositado nas centrais de bioinformática do Projeto *Genolyptus*, incluindo 21.442 sequências únicas. É importante salientar que dentre as ESTs depositadas aquelas derivadas das bibliotecas de flores e botões florais de *E. grandis* não chegaram a 700 sequências válidas (G. Pasquali, comunicação pessoal).

Muitos estudos nas últimas décadas foram realizados com o objetivo de estudar a fisiologia molecular da formação das flores e do processo de florescimento. Entretanto, não há muitos estudos moleculares sobre a expressão gênica em botões florais em antese e flores abertas. Aliado a este fato, à importância sócio-econômica que a silvicultura desempenha no mercado brasileiro e ao aumento progressivo de áreas plantadas com florestas de eucalipto, por meio deste trabalho foi proposta a obtenção de um número maior de ESTs e o levantamento de genes expressos em flores abertas e botões florais em antese de *E. grandis*, a fim de compreender o papel destes genes na abertura da flor e na manutenção da estrutura floral.

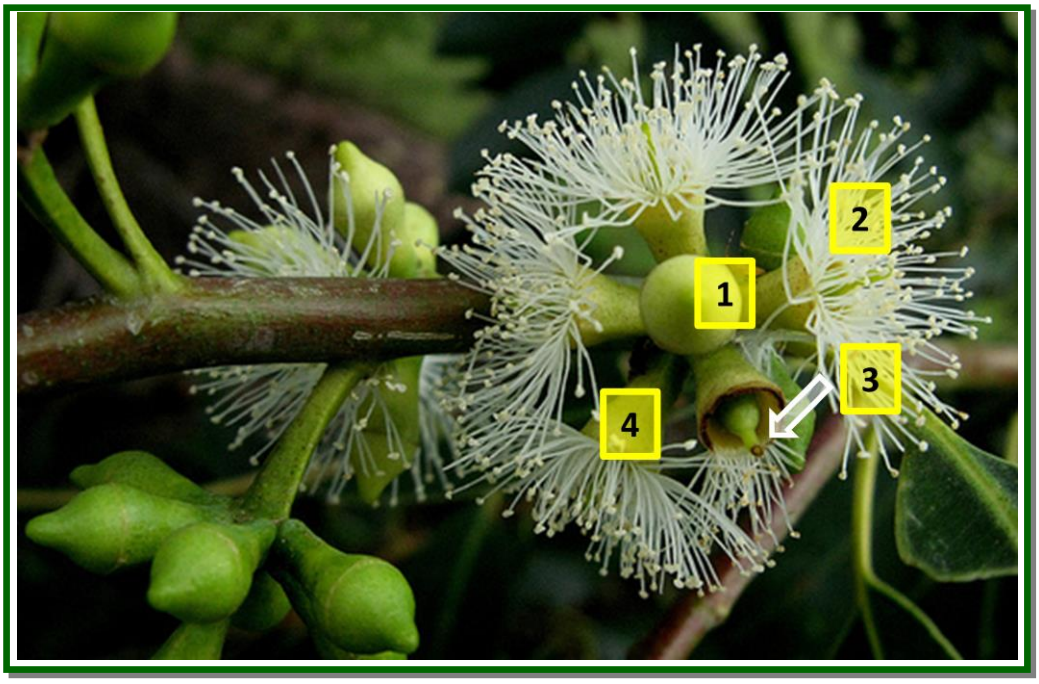


Figura 3 – **Inflorescência de *Eucalyptus grandis* com flores e botões florais em diferentes estádios de florescimento.** As estruturas destacadas com números correspondem àquelas utilizadas para a construção das bibliotecas de expressão. 1 – Opérculo (pétalas/sépalas); 2 – Anteras/estames; 3 – Pistilo; e 4 – Carpelos/receptáculo floral.

Foto: http://www.flickr.com/photos/9764939@N08/2281604435/

Resultados e Discussão

Sequenciamento de cDNAs e Validação de ESTs de Flores de *E. grandis*

Com vistas a gerar uma coleção de ESTs derivadas de flores de *E. grandis*, uma série de tentativas de construção de bibliotecas de cDNA de flores inteiras em diferentes estádios do desenvolvimento foram realizadas, sem, no entanto, obter-se sucesso. As bibliotecas produzidas a partir de flores inteiras, à semelhança do que ocorreu durante o Projeto Genolyptus (G. Pasquali, comunicação pessoal), apresentaram títulos muito baixos, isto é, número muito baixo de colônias de *Escherichia coli* contendo plasmídeos recombinantes com insertos de cDNA. Embora as preparações de RNA e mRNA fossem satisfatórias, bons títulos de bibliotecas nunca foram obtidos, mesmo com o emprego de diferentes sistemas de síntese de bibliotecas de cDNA (resultados não apresentados). Por esta razão, decidiu-se realizar extrações de RNA total a partir de diferentes partes das flores de *E. grandis*, conforme ilustradas na Figura 3. As flores e os botões florais foram coletados de três indivíduos distintos de *E. grandis* com quatro anos de idade. Os órgãos foram seccionados com bisturi e os conjuntos de suas quatro partes foram imediatamente congelados em nitrogênio líquido. As partes separadas das flores e botões foram designadas por (i) opérculo (pétalas/sépalas), (ii) anteras/estames, (iii) pistilo e (iv) carpelos/receptáculo floral. Após a extração de RNA total, a síntese de cDNA foi realizada utilizando-se o *kit SuperScript III* (Invitrogen). Os clones de cDNA foram subsequentemente ligados ao plasmídeo pGEM-T-Easy (Promega) e clonados em *E. coli* DH10B. Após a preparação de plasmídeos em microplacas de 96 poços por lise alcalina seguida da filtração em microplacas filtrantes, os insertos foram sequenciados em sistema *ABI Prism 3100 Genetic Analyzer* (Applied Biosystems) a partir dos *primers* universais *M13 -21 forward* e *T7*.

Um total de 3.190 ESTs foram geradas ao todo, sendo validadas apenas aquelas com 150 pb de comprimento mínimo. Assim, após exclusão de ESTs mais curtas, remoção de regiões de baixa qualidade, sequências do vetor e possíveis artefatos, 1.641 ESTs foram consideradas válidas, ou seja 51,4% do total. Conforme indicado na Tabela 2, utilizando o sequenciador ABI 3130 (Applied Biosystems), Vinod *et al*. [56] geraram ESTs de flores masculinas de *Pandanus fascicularis,* obtendo 24,4% de sequências válidas, e Moccia *et* al. [57] analisaram ESTs de *Silene latifolia* com a finalidade de selecionar marcadores de sequências simples repetitivas (do inglês, *Simple Sequence Repeat*, SSR) e obtiveram 82,9% de sequências válidas. Beulé *et al*. [58] analisaram ESTs geradas a partir do sequenciamento de inflorescências imaturas de *Elaeis guinensis*, obtendo resultados semelhantes aos encontrados em flores de *Silene latifolia*, com 83,0% de ESTs válidas. Provavelmente, a maior eficiência na validação das ESTs foi resultante da purificação prévia de mRNA realizada pelos autores citados para a construção das bibliotecas.

Guo *et al*. [59] compararam o transcriptoma de flores de diferentes sexos de *Cucumis sativus* utilizando a técnica de pirosequenciamento massivo pela plataforma *Roche 454* e obtiveram 87,4% das ESTs válidas (Tabela 2). As diferenças percentuais encontradas entre os valores de ESTs válidas obtidas no sequenciamento automático convencional (baseado no método de Sänger) e o pirosequenciamento massivo deve-se, principalmente, ao tamanho das sequências geradas. Enquanto o sequenciamento convencional permite gerar sequências de 50 a 1.200 pb, o pirosequenciamento gera sequências menores, com no máximo 300 pb, porém em números na ordem de centenas de milhares. Portanto, este último método permite uma cobertura maior de sequenciamento e, também, uma grande redundância de ESTs, o que é vantajoso para estudos de expressão gênica. De acordo com os resultados obtidos em flores de outras espécies utilizando técnicas de sequenciamento convencional, como apresentado na Tabela 2, a percentagem de ESTs válidas geradas para flores de *E. grandis* no presente trabalho é significativamente representativa para este órgão da planta, conforme encontrado na literatura.

Tabela 2 – Quadro comparativo de resultados do sequenciamento automático de cDNAs de flores e botões florais de nove espécies de plantas, incluindo *E. grandis*.

	Gerbera hybrida	***Phanalaenopsis equestris***	***Petunia x hybrida***	***Pandanus fascicularis***	***Silene latifolia***	***Cucumis sativus***	***Elaeis guinensis***	***Panax quinquefolius***	***Eucalyptus grandis***
Estrutura analisada	Flores	Botões florais	Flores, pólen e tubos polínicos	Flores masculinas	Flores	Flores masculinas e femininas	Inflorescências imaturas	Flores	Flores
Metodologia	ABI 3700	ABI 377	ABI 377	ABI 3130	ABI 3130	Roche 454	N.I.	ABI 3730	ABI 3700
Total de ESTs	N.I.	N.I.	N.I.	4.000	4.416	405.000	1.920	1.440	3.190
ESTs válidas	16.994	5.593	5.043	977 (24,4%)	3.662 (82,9%)	353.941 (87,4%)	1.594 (83,0%)	1.196 (83,0%)	1.641 (51,4%)
Unigenes	8.098 (47,6%)	3.688 (65,9%)	3.021 (65,9%)	511 (53,3%)	3.105 (84,8%)	81.401 (23,0%)	1.350 (84,7%)	387 (32,3%)	403 (24,5%)
Clusters	2.591 (32%)	732 (20%)	661 (20%)	83 (16%)	432 (13,9%)	28.453 (35%)	145 (11%)	77 (19%)	120 (30%)
Singletons	5.507 (68%)	2.956 (80%)	2.410 (80%)	428 (84%)	2.673 (86,1%)	52.948 (65%)	1.204 (88%)	310 (81%)	283 (70%)
Referência	Laitinen *et al*. [60]	Tsai *et al*. [61]	Shimamura *et al.* [58]	Vinod *et al*. [56]	Moccia *et al.* [57]	Guo *et al*. [59]	Beulé *et al*. [62]	Wu *et al*. [63]	Este trabalho

(N.I., Não Informado.)

As ESTs de alta qualidade apresentaram um comprimento médio de 525 pb. Apesar de um número significativo de ESTs possuir um comprimento pequeno, inferior a 150 pb, ou uma baixa qualidade de sequenciamento, mais de 65% apresentaram comprimento maior ou igual à média. As ESTs válidas distribuiram-se da seguinte forma entre as bibliotecas: 398 ESTs derivadas da biblioteca de sépalas/pétalas; 454 da biblioteca de anteras/estames; 397 da biblioteca de pistilos; e 392 ESTs da biblioteca de carpelos/receptáculo floral (Tabela 3).

As ESTs geradas por este estudo foram submetidas à montagem de agrupamentos (*clustering*) utilizando-se o *software CodonCode Aligner* (http://www.codoncode.com/aligner/index.htm). Um total de 403 unigenes foi obtido, dentre os quais 120 *clusters* constituídos por 1.358 ESTs (30%) e 283 *singletons* (70%), apresentando um comprimento médio de 870 pb e 316 pb, respectivamente. Conforme pode ser observado na Tabela 2 acima, estes resultados estão de acordo com os trabalhos de Laitinen *et al*. [60], que obteve 32% dos unigenes agrupados em *clusters* e 68% em *singletons*; Tsai *et al*. [61] e Shimamura *et al*. [58], com aproximadamente 20% das ESTs agrupadas em *clusters* e 80% em *singletons*; Moccia *et al*. [57] obtiveram aproximadamente 14% das ESTs agrupadas em *clusters* e 86% em *singletons*; Beulé *et al.*, obtiveram 11% agrupadas em *clusters* e 88% em *singletons*; e Vinod *et al*. [56], que obtiveram aproximadamente 16% de unigenes agrupados em *clusters* e 84% em *singletons*. Os resultados obtidos por Guo *et al*. [59] foram muito semelhantes ao encontrado para as análises de *E. grandis*, agrupando 35% das ESTs em *clusters* e 65% em *singletons*, embora a metodologia de sequenciamento tenha sido distinta.

A distribuição dos *singletons*, *clusters* e unigenes conforme a anotação conduzida com o software Blast2GO [64] estão representados na porção inferior da Tabela 3, revelando que 37% dos unigenes foram anotados como proteínas de função conhecida, enquanto 18,5% foram anotadas como proteínas de função desconhecida e 44,4% não apresentaram similaridade com proteínas depositadas em bancos de dados. Novamente, uma distribuição semelhante na anotação dos dados foi obtida por Laitinen *et al*. [60], onde 29,2% dos unigenes foram caracterizados com proteínas de função conhecida, 12,4% foram anotadas com proteínas de função desconhecida e

58,8% não apresentaram similaridade. Shimamura *et al*. [58] obtiveram 60% dos unigenes anotados como proteínas de função conhecida e 40% com função desconhecida.

Tabela 3 – Total de ESTs geradas e validadas a partir de flores de *E. grandis*, número de *clusters* e *singletons* e anotação das sequências obtidas para cada biblioteca de cDNA segundo o programa Blast2GO.

		Partes das Flores				
		Pétalas/ Sépalas	**Anteras/ Estames**	**Pistilo**	**Carpelos/ Receptáculo floral**	**Total**
ESTs	Total	576	876	804	934	3.190
	Válidas	398	454	397	392	1.641
	Singletons	64	86	61	72	283
	Clusters	5	50	7	58*	120
Anotação	N° total de Unigenes	69	136	68	130**	403
	Proteínas de função conhecida	12	59	6	72	149
	Proteínas de função desconhecida	11	25	9	21	75
	Nenhuma homologia (*no hits found*)	52	45	47	35	179

* Valor correspondente a 41 *clusters* exclusivos da biblioteca de carpelos/receptáculo floral e 17 *clusters* compartilhados entre as bibliotecas de pétalas/sépalas, anteras/estames e pistilo.

** Inclui *clusters* compartilhados com as bibliotecas de pétalas/sépalas, anteras/estames e pistilo.

Um diagrama de Venn foi construído para representar o compartilhamento de ESTs entre as bibliotecas. Conforme ilustrado pela Figura 4, as intersecções representam *clusters* cujas sequências constituintes foram comuns entre bibliotecas. Desta forma, as sequências estão assim distribuídas entre as bibliotecas: 69 unigenes estão presentes na biblioteca de sépalas/pétalas, 136 em anteras/estames, 68 em pistilos e 130 em carpelos/receptáculo floral. Curiosamente, apenas a biblioteca de carpelos/receptáculo floral compartilhou sequências com as demais bibliotecas, totalizando 17 unigenes. Destas sequências, 3,22% das ESTs (13 unigenes) foram observadas nas bibliotecas de anteras/estames e carpelos/receptáculo floral. Em contraste, até 32,3% dos unigenes foram específicos à biblioteca de anteras/estames e 28% à biblioteca de carpelos/receptáculos

florais. Os unigenes comuns encontrados entre as bibliotecas de carpelos/receptáculo floral e pistilos totalizaram 0,74% das sequências (3 unigenes), e 0,25% (1 unigene) entre as bibliotecas de carpelo/receptáculo floral e sépalas/pétalas. Os números de unigenes exclusivos encontrados nas bibliotecas de pistilo e sépalas/pétalas foram próximos entre si, com 68 e 69 unigenes, respectivamente. A biblioteca de carpelos/receptáculo floral apresentou 112 unigenes exclusivos.

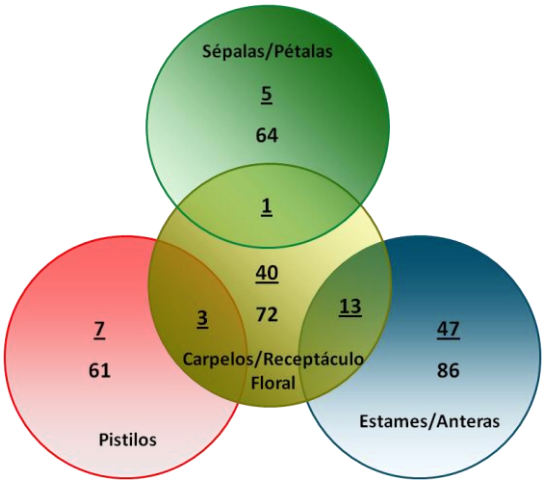


Figura 4 – Diagrama de Venn ilustrando a sobreposição de sequências derivadas de genes expressos encontrados em bibliotecas de pétalas/sépalas, anteras/estames, pistilos e carpelos/receptáculo floral de *E. grandis*. Os números em regiões sobrepostas indicam as ESTs compartilhadas entre os respectivos conjuntos. Os números em regiões não-sobrepostas indicam ESTs biblioteca-específicas. Os valores grifados correspondem a *clusters* e os não grifados correspondem a *singletons*.

Ontologia Gênica do Transcriptoma de Flores de *E. grandis*

Para inferir possíveis funções dos prováveis genes obtidos pelo sequenciamento de cDNAs e geração de ESTs de flores de *E. grandis*, uma anotação eletrônica foi realizada com base em alinhamentos dos unigenes, comparando-se estas sequências contra os bancos de dados GenBank (http://blast.ncbi.nlm.nih.gov/Blast.cgi) e TAIR (http://www.arabidopsis.org/Blast/). Para tanto, utilizou-se o programa *Basic Local Alignment Search Tool* (BLAST) [65] disponibilizado pelo *National Center for*

Biotechnology Information (NCBI), selecionando-se a opção de proteínas não-redundantes (nr), com um valor estimado (*e-value*) de corte de 1×10^{-5}.

Uma subsequente análise dos unigenes das flores de *E. grandis* para fins de classificação segundo a ontologia gênica (GO, do inglês, *Gene Ontology*) foi realizada empregando-se novamente o *software* Blast2GO [64]. Este *software* permite classificar os unigenes segundo a semelhança das sequências peptídicas deduzidas com domínios protéicos conservados e anotados nos bancos de dados *InterPro* [66] e *Pfam* [67]. Para apenas um total de 62 dos unigenes (26%) foi designada (pelo menos) uma categoria de GO. Destas, apenas a 0,76% das sequências foi atribuída uma categoria em processo biológico, 21,05% em função molecular e 7,86% em componentes celulares. Na Figura 5 está apresentada a classificação funcional obtida para os unigenes nas respectivas categorias.

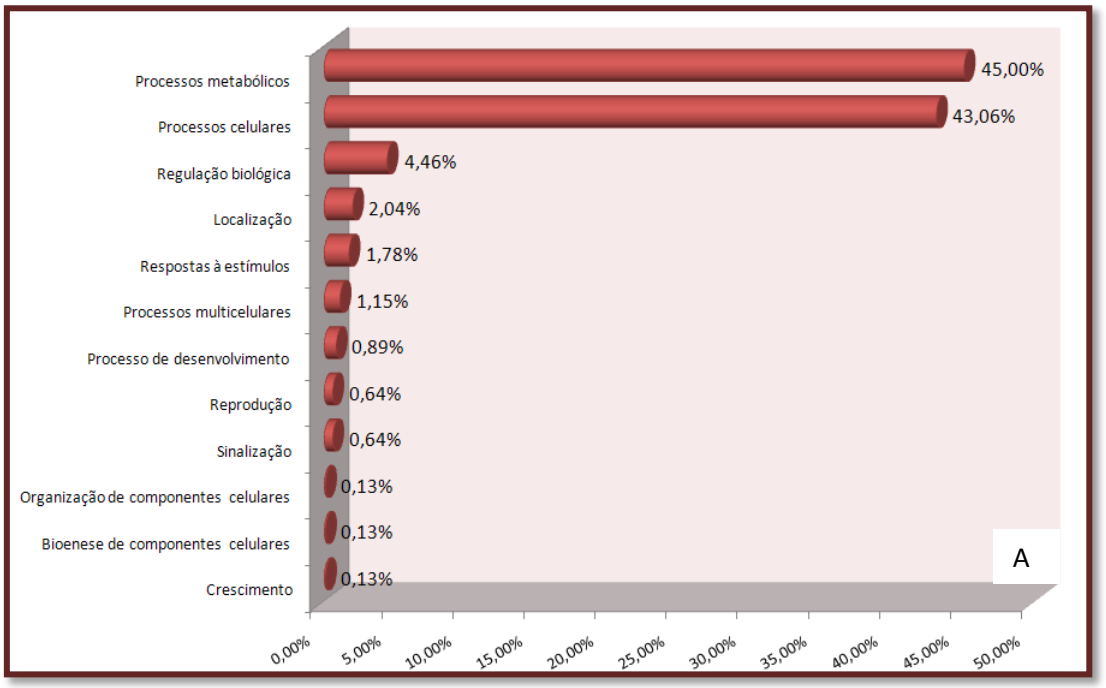
Conforme pode ser observado na Figura 5A, processos metabólicos, processos celulares e regulação biológica foram os grupos mais bem representados pelos 62 unigenes classificados, com 45, 43 e 4,5% de representatividade, respectivamente, para a categoria de processos biológicos. Nos trabalhos conduzidos por Tsai *et al*. [61] e Laitinen *et al*. [60], o grupo mais abundande também foi o de genes envolvidos em processos metabólicos, cruciais para a manutenção da homeostase da flor. Genes envolvidos em processos celulares foram o segundo grupo mais representado no trabalho de Laitinen *et al*. [60], e o terceiro grupo nas análises realizadas por Vinod *et al*. [56].

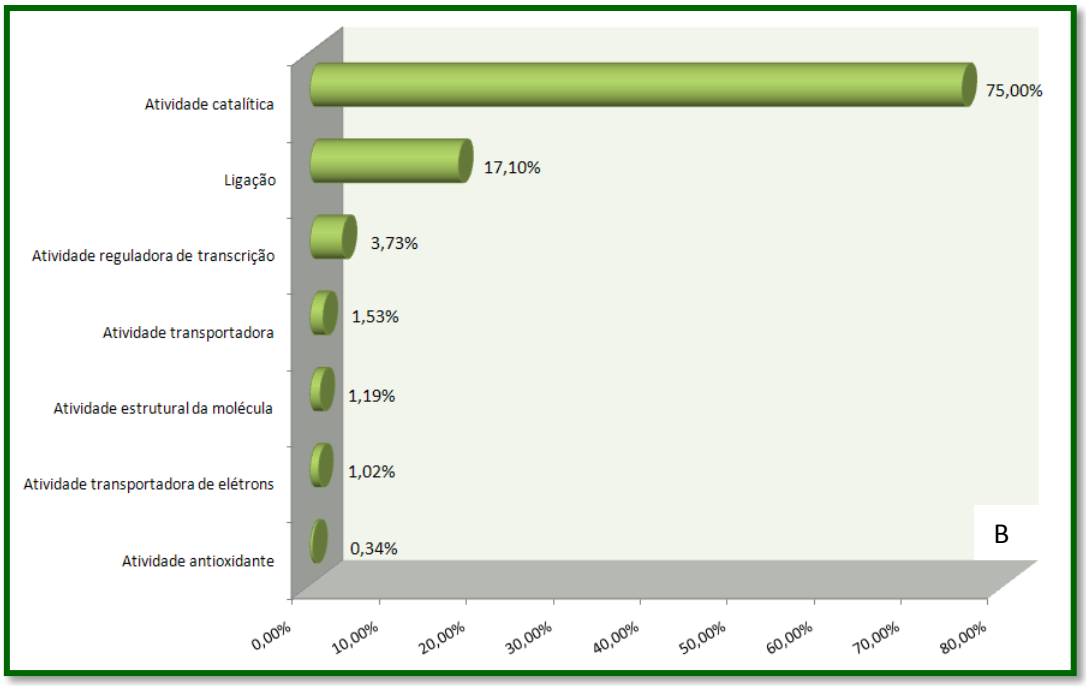
Os grupos relacionados à atividade catalítica (75,0%), à ligação (17,1%), à atividade reguladora de transcrição (3,73%) e à atividade transportadora (1,53%) reuniram o maior número de unigenes de flores de *E. grandis* pela categorização GO em função molecular (Figura 5B). Nas análises de ESTs de flores de *S. latifolia* [57], os grupos mais abundantes foram os relacionados à ligação e à atividade catalítica, seguido pelo grupo de ligação a ácidos nucléicos e a proteínas. Nesta espécie, o grupo representando genes envolvidos na atividade reguladora de transcrição é bem pouco representativo, sendo classificado como um dos três grupos menos representados. A avaliação do perfil transcricional de flores de *P. quinquefolius* realizada por Wu *et al*.

[63] também mostrou que os grupos de ligação e de atividade catalítica são os mais que possuem maior representatividade, assim como observado para os unigenes das flores de *E. grandis*, seguidos pelos grupos de estrutura de moléculas e atividade reguladora de transcrição.

Com relação à categoria GO de componentes celulares em que os unigenes de flores de *E. grandis* foram agrupados, mais de 80% classificaram-se nas categorias de células e organelas. Resultado semelhante foi obtido por Moccia *et al*. [57], onde genes classificados nos grupos relacionados à célula (intracelular, citoplasma e membrana) e a organelas (plastídeos e mitocôndrias) foram os mais bem representados.

No presente trabalho, foram adicionalmente anotadas várias ESTs com funções relacionadas a processos metabólicos e celulares, senescência, defesa e resposta a estímulos ambientais e hormonais, confirmando um panorama de manutenção celular e resposta à cronologia da floração, preparando a flor para o momento da fertilização e do início da formação do fruto.





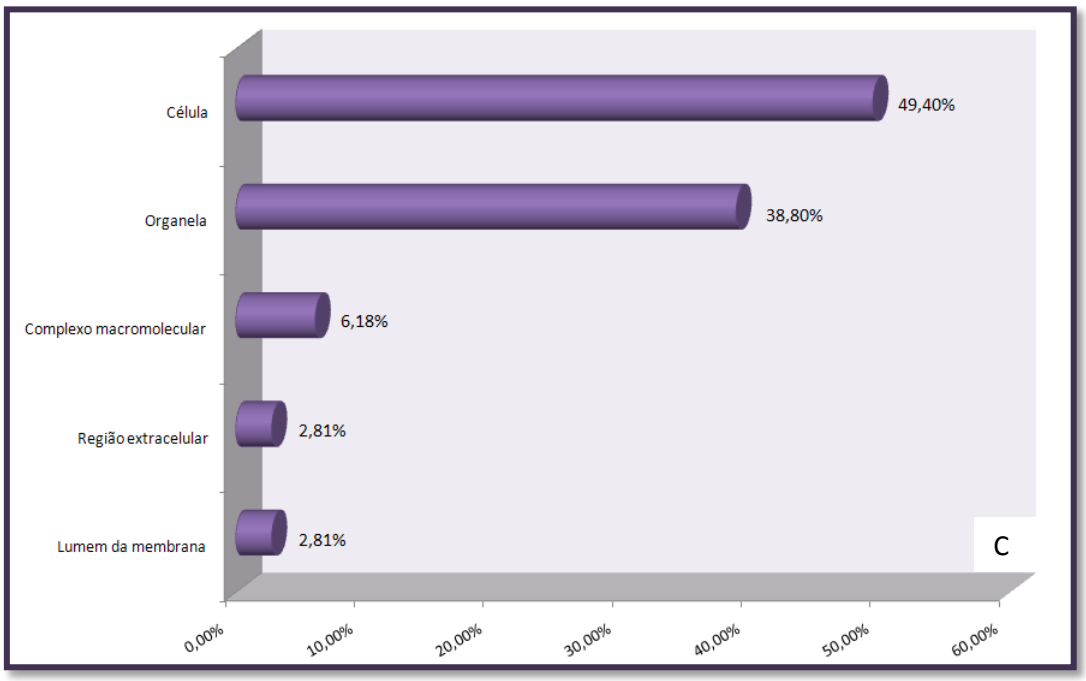


Figura 5 – Classificação de 62 unigenes de flores e botões florais de *E. grandis* conforme a ontologia gênica atribuída por Blast2GO nas categorias de processos biológicos (A), função molecular (B) e componentes celulares (C). As sequências com valores de *e-value* inferiores a 1 x 10^{-6} foram anotadas com tBlastX e designadas para a categoria funcional com base nos resultados de escore de Blast mais elevados. As percentagens são as proporções para o conjunto total de unigenes (62), exceto àqueles não anotados.

Principais Classes de Genes Expressos em Flores de *E. grandis*

As células especializam-se em funções fisiológicas distintas, determinadas pelo repertório de proteínas associado a cada tipo de metabolismo. Este repertório de proteínas é determinado, por sua vez, pela regulação da expressão dos genes, especialmente em nível da transcrição. As proteínas com funções gerais, como aquelas envolvidas no metabolismo energético, na constituição do citoesqueleto, os polipeptídeos ribossomais, as histonas, entre outras, são codificadas por genes mantenedores da célula e geralmente constitutivos (*housekeeping genes*), os quais são transcritos em todos os tipos celulares. Outros grupos de proteínas acumulam-se nas células em resposta a sinais fisiológicos específicos, determinados por reguladores do crescimento, estresses bióticos ou abióticos ou pela disponibilidade de nutrientes. Após a classificação das ESTs em suas categorias funcionais segundo a GO, uma melhor caracterização das sequências expressas em flores e botões florais de *E. grandis* foi realizada segundo dois aspectos principais: (i) *clusters* que possuem maior representatividadepor ESTs, e (ii) fatores de transcrição, em virtude da relevância destas últimas proteínas na determinação das identidades tissulares e funções fisiológicas da flor. Os *clusters* que possuem maior representatividadeencontrados estão indicados na Tabela 4, e estes estão divididos essencialmente em dois grupos funcionais: aqueles relacionados à manutenção celular, e aqueles dedicados à senescência. Assim, o maior número de ESTs identificadas referem-se a genes de expressão constitutiva ou a genes potencialmente relacionados aos sinais fisiológicos de senescência da flor.

Tabela 4 - **Anotação *in silico* por BlastX frente aos bancos TAIR e GenBank (NCBI) de *clusters* obtidos a partir de ESTs válidas das bibliotecas de flores de *E. grandis*.** Os *clusters* estão representados em ordem decrescente de suas composições (número de ESTs ou *reads*), conforme indicado na última coluna. As linhas destacadas em verde representam genes cuja expressão foi avaliada por RT-qPCR.

Cluster	TAIR			NCBI			Partes da Flor*	N° *Reads*
	ID	Anotação (Melhor *hit* no BlastX)	e-value	ID	Anotação (Melhor *hit* no BlastX)	e-value		
001	AT1G07600.1	Symbols: MT1A, MT1, LSR4, MT-Q, ATMT-2, ATMT-Q, ATMT1 | metallothionein 1A	3e-23	gb|AAF79580.1|	F22G5.2 [*A.thaliana*]	1e-07	S/P	253
002	AT4G38770.1	Symbols: PRP4, ATPRP4 | proline-rich protein 4	1e-13	emb|CAA04449.1|	proline-rich protein [*Solanum tuberosum*]	4e-29	P	162
003	AT2G38940.1	ATPT2 (PHOSPHATE TRANSPORTER 2); carbohydrate transmembrane transporter/ phosphatetransmembrane transporter/ sugar:hydrogen ion symporter	2e-58	dbj|BAE94387.1|	putative phosphate transporter [*E. camaldulensis*]	1e-66	A/E	42
004	AT1G30220.1	Symbols: ATINT2, INT2 | inositol transporter 2	9e-20	gb|ACA04864.1|	carbohydrate transporter/sugar porter [Picea abies]	9e-20	A/E C/RF	41

Cluster	TAIR			NCBI			Partes da Flor*	N° *Reads*
	ID	**Anotação (Melhor *hit* no BlastX)**	**e-value**	**ID**	**Anotação (Melhor *hit* no BlastX)**	**e-value**		
005	AT5G41870.1	Pectin lyase-like superfamily protein	7e-19	ref|XP_002509949.1	Polygalacturonase precursor, putative [*Ricinus communis*]	4e-20	A/E C/RF	40
006	AT3G24503.1	ALDH2C4, ALDH1A, REF1 | aldehyde dehydrogenase 2C4	1e-50	ref|XP_002514330.1	aldehyde dehydrogenase, putative [*R.communis*]	2e-58	A/E C/RF	35
007		No hit found			No hit found		A/E C/RF	33
008	AT4G02080.1	ASAR1 (*A. thaliana* secretion-associated RAS super family 2); GTP binding	9e-17	ref|NP_192117.1|	ASAR1 (*A.thaliana* secretion-associated RAS super family 2); GTP binding	1e-14	A/E	31
009	AT1G12310.1	calmodulin, putative	1e-22	ref|XP_002304435.1	predicted protein [*Populus trichocarpa*]	1e-22	C/RF	31
010	AT1G53920.1	GLIP5 | GDSL-motif lipase 5	2e-19	ref|XP_002533919.1	zinc finger protein, putative [*R.communis]*	9e-25	C/RF	29
011		No hit found			No hit found		A/E C/RF	25
018	AT5G40990.1	GLIP1 GDSL lipase 1	6e-19	ref|XP_002533919.1	zinc finger protein, putative [*R.communis*]	6e-25	C/RF	20
023	AT5G58860.1	CYP86A1 (cytochrome P450, family 86, subfamily A, polypeptide 1); oxygen binding	3e-23	emb|CAA62082.1|	cytochrome p450 [*A.thaliana*]	4e-21	P C/RF	18

Cluster	TAIR			NCBI			Partes da Flor*	N° *Reads*
	ID	Anotação (Melhor *hit* no BlastX)	e-value	ID	Anotação (Melhor *hit* no BlastX)	e-value		
024	AT5G16940.2	carbon-sulfur lyases	7e-41	ref|XP_002302935.1	predicted protein [*P. trichocarpa*]	6e-44	C/RF	18
030	AT2G35680.1	Phosphotyrosine protein phosphatases superfamily Protein	6e-85	ref|XP_002514193.1|	Protein-tyrosine phosphatase mitochondrial 1, mitochondrial precursor, putative [*R. communis*]	3e-98	S/P	15
031	AT4G19420.1	Pectinacetylesterase family protein	6e-19	gb|ACF05806.1	PAE [*Litchi chinensis*]	2e-23	A/E	15
033	ATCG01010.1	Chloroplast encoded NADH dehydrogenase unit	4e-36	gb|AAF76341.1|	NADH dehydrogenase subunit F [*Rhynchocalyx lawsonioides*]	2e-35	A/E	14
034	AT2G47140.1	NAD(P)-binding Rossmann-fold superfamily protein	8e-64	ref|XP_002511512.1	short chain alcohol dehydrogenase, putative [*R. communis*]	1e-78	A/E C/RF	14
035	AT1G79930.2	HSP91 | heat shock protein 91	1e-29	emb|CAA94389.1	heat-shock protein [*A.thaliana*]	2e-27	A/E C/RF	14
036	AT3G10160.1	ATDFC, DFC, FPGS2 | DHFS-FPGS homolog C	8e-21	ref|XP_002322647.1	folylpolyglutamate synthase[*P. trichocarpa*]	2e-22	C/RF	14
037	AT3G44290.1	Anac060, NAC060 NAC domain containing protein 60	8e-12	ref|XP_002313601.1	NAC domain protein, IPR003441 [*P.trichocarpa*]	4e-10	C/RF	14

Cluster	TAIR			NCBI			Partes da Flor*	N° *Reads*
	ID	Anotação (Melhor *hit* no BlastX)	e-value	ID	Anotação (Melhor *hit* no BlastX)	e-value		
038	AT1G12520.1	ATCCS, CCS1 | CCS1 (copper chaperone for superoxide dismutase 1); superoxide dismutase copper chaperone	4e-07	gb|AAK01931.1|AF329816_1	Cu/Zn-superoxide dismutase copper chaperone precursor [*Glycine max*]	8e-08	C/RF	14
042	AT5G20720.2	CPN20 (CHAPERONIN 20); calmodulin binding	6e-08	gb|AAT80888.1|	chloroplast chaperonin 21 [*Vitis vinifera*]	2e-06	C/RF	11
044		No hit found		dbj|BAB33421.1	putative senescence-associated protein [*Pisum sativum*]	4e-68	S/P C/RF	10
045	AT5G37600.1	ATGSR1 | ATGSR1 (Arabidopsis thaliana glutamine synthase clone R1); glutamate-ammonia ligase	6e-14	gb|AAT39510.1	glutamine synthetase [*Elaeagnus umbellata*]	3e-13	C/RF	10
046	AT1G55020.1	LOX1 (Lipoxygenase 1); lipoxygenase	2e-15	emb|CAB83038.1|	lipoxygenase-9 [*Cucumis sativus*]	5e-17	A/E	10
047	AT2G27300.1	ANAC040, NTL8 NTM1-like 8	7e-11	ref|XP_002313601.1	NAC domain protein, IPR003441 [*P.trichocarpa*]	3e-09	C/RF	9
048	AT5G57560.1	TCH4, XTH22 Xyloglucan endotransglucosylase/hydrolase family protein	5e-40	ref|XP_002270416.1	PREDICTED: hypothetical protein [*V.vinifera*]	6e-45	A/E	8

Cluster	TAIR			NCBI			Partes da Flor*	N° *Reads*
	ID	Anotação (Melhor *hit* no BlastX)	e-value	ID	Anotação (Melhor *hit* no BlastX)	e-value		
049	AT4G16720.1	60S ribosomal protein L15 (RPL15A)	5e-45	gb|AAT85124.1|	putative 60s ribosomal protein L15 [*Oryza sativa* (japonica)]	2e-44	C/RF	8
053	AT4G37530.1	peroxidase, putative	6e-12	dbj|BAC16194.1|	putative peroxidase [*O.sativa* Japonica]	9e-13	C/RF	7
055	AT3G16080.1	60S ribosomal protein L37 (RPL37C)	7e-09	gb|AAS47512.1|	ribosomal protein L37 [*G. max*]	1e-07	C/RF	7
067	AT5G23860.1	TUB8 (tubulin beta-8)	4e-22	gb|ABY86655.1|	beta-tubulin 4 [*Gossypium hirsutum*]	4e-20	A/E C/RF	6
068	AT3G54400.1	aspartyl protease family protein	4e-08	gb|ACE96805.1|	aspartyl protease [*P. tremula*]	9e-14	C/RF	6
069	AT1G20010.1	TUB5 (tubulin beta-5 chain)	3e-06	dbj|BAE47136.1|	beta tubulin like protein [*Pyrus pyrifolia* var. culta]	4e-06	A/E C/RF	6
070	AT3G47950.1	AHA4 (Arabidopsis H(+)-ATPase 4)	2e-17	dbj|BAC77533.1|	plasma membrane H+-ATPase [*Sesbania rostrata*]	7e-17	A/E	5
071	AT2G32670.1	ATVAMP725 (Arabidopsis thaliana vesicle-associated membrane protein 725)	3e-17	ref|NP_180826.2	ATVAMP725 (*A.thaliana* vesicle-associated membrane protein 725)	3e-15	A/E	5
072	AT5G08420.1	RNA binding	9e-18	gb|AAP54059.1|	Ribosomal RNA assembly protein mis3, putative, expressed [*O. sativa* (japonica)]	2e-20	C/RF	5

Cluster	TAIR			NCBI			Partes da Flor*	N° *Reads*
	ID	Anotação (Melhor *hit* no BlastX)	e-value	ID	Anotação (Melhor *hit* no BlastX)	e-value		
073	AT4G00570.1	malate oxidoreductase, putative	7e-07	gb|ABR13310.1|	putative NAD dependent malic enzyme [*Prunus dulcis*]	4e-05	C/RF	5
074	AT1G26880.1	60S ribosomal protein L34 (RPL34A)	2e-47	gb|AAA86953.1|	60S ribosomal protein L34 [*A. thaliana*]	8e-47	C/RF	5
075	AT4G35090.2	CAT2 (CATALASE 2); catalase	5e-45	emb|CAB16749.1|	catalase [*Soldanella alpina*]	8e-44	C/RF	5
082	AT3G51840.1	ACX4 (ACYL-COA OXIDASE 4); oxidoreductase	2e-26	dbj|BAG09369.1|	peroxisomal acyl-CoA oxidase [*G .max*]	2e-25	C/RF	4
083	AT1G09750.1	chloroplast nucleoid DNA-binding protein-related	5e-31	ref|NP_563851.1|	chloroplast nucleoid DNA-binding protein-related [*A. thaliana*]	7e-29	C/RF	4
084	AT5G54840.2	GTP-binding family protein	9e-12	gb|AAM64604.1|	putative SGP1 monomeric G-protein [*A. thaliana*]	9e-10	C/RF	4
087	AT2G34520.1	RPS14 (MITOCHONDRIAL RIBOSOMAL PROTEIN S14); structural constituent of ribosome	5e-24	dbj|BAD83430.1|	ribosomal protein S14 [*Nicotiana tabacum*]	2e-26	C/RF	4
090	AT3G62250.1	UBQ5 (UBIQUITIN 5); protein binding	1e-11	gb|ABN72579.1|	ubiquitin extension protein [*Hevea brasiliensis*]	1e-22	C/RF	3
096	AT5G24090.1	acidic endochitinase (CHIB1)	4e-16	gb|ABN03967.1|	acidic chitinase [Gossypium hirsutum]	5e-16	C/RF	2

Cluster	TAIR			NCBI			Partes da Flor*	N° *Reads*
	ID	Anotação (Melhor *hit* no BlastX)	e-value	ID	Anotação (Melhor *hit* no BlastX)	e-value		
106	AT4G34410.1	AP2 domain-containing transcription factor	6e-24	dbj|BAG50061.1	transcription factor AP2-EREBP [*Lotus japonicus*]	8e-25	C/RF	2
110	AT1G24260.3	SEP3| K-box region and MADS-box transcription factor	1e-04	gb|AAN15182.1	MADS box protein GHMADS-1 [*G.hirsutum*]	2e-10	P	2
111	AT3G19184.1	AP2/B3-like transcriptional factor family protein	2e-85	ref|XP_002282080.1	hypothetical protein [*V.vinifera*]	2e-62	C/RF	2
118	AT5G21160.2	LA RNA-binding protein	6e-65	emb|CBI24386.3	unnamed protein product [*V.vinifera*]	2e-64	C/RF	2
	AT2G25930.1	ELF3, PYK20 | hydroxyproline-rich glycoprotein family protein	4e-21	ref|XP_002278577.1	hypothetical protein [*V. vinifera*]	4e-57	P	1

* Partes da Flor: S/P – pétalas/sépalas; A/E – anteras/estames; P – pistilos; C/RF – carpelos/receptáculos florais.

Aos 50 *clusters* representados na Tabela 4 foram também atribuídas identidades *in silico* com mais alta robustez frente aos bancos de dados NCBI e TAIR utilizando-se o programa BlastX. Todos os unigenes indicados na Tabela 4 pertencem aos três grupos mais bem representados na categoria GO de processos metabólicos, conforme Figura 5A.

Algumas sequências relacionadas ao transporte de íons foram encontradas nas bibliotecas de flores de *E. grandis*. Dentre elas, as sequências anotadas como metalotioneínas (*Cluster 001*), calmodulinas (*Cluster 009*), citocromos P_{450} (*Cluster 023*) e proteínas RING do tipo "dedo-de-zinco" (Cluster 010). A sequência consenso anotada como metalotioneína representa o *cluster* mais populado de toda a análise, e sua presença foi detectada na biblioteca de pétalas/sépalas. Metalotioneínas são proteínas ubíquas de baixo peso molecular, com 60 a 80 resíduos de aminoácidos, alto teor de metais, ricas em cisteínas e encontradas em animais, plantas e fungos. Sua função está relacionada ao sequestro de íons metálicos de zinco, cobre, cádmio e mercúrio [68-70]. Estas proteínas exercem influência sobre diversos processos celulares incluindo reações de óxido-redução, reações catalisadas por enzimas e metabolismo de ácidos nucléicos. Os genes que codificam estas proteínas mostram-se diferencialmente expressos em plantas e são induzidos por uma série de estímulos como ozônio [71], choque térmico, metais pesados e escassez de sacarose [72]. Os genes de metalotioneínas também são induzidos durante o desenvolvimento das plantas como, por exemplo, na formação do *tapetum* de milho [68] e dos tecidos de maturação de frutos como uva [69, 73], maçã [74] e framboesa [75], e no processo de senescência de folhas de *Brassica napus* [76, 77].

Vários estudos permitiram relacionar a expressão de genes de metalotioneínas ao processo de senescência das pétalas, com níveis variados de expressão. Breeze *et al.* [78] observaram que 19% das ESTs encontradas nas pétalas de *Alstroemeria* codificavam esta classe de proteínas, enquanto Channeliere *et al*. [79] encontraram-nas com 11% de representatividade em pétalas de rosas. Tsai *et al*. [61] avaliaram a expressão gênica global em flores de *P. equestris* e encontraram mRNA para metalotioneína como o sétimo transcrito mais bem representado nesta análise. Hunter *et al*. [69] encontraram mRNAs para metalotioneínas presentes entre os

transcritos associados à senescência de *Narcisus pseudonarcisus*. Portanto, a presença de transcritos para metalotioneínas em altos níveis na biblioteca de pétalas/sépalas de *E. grandis* não é inesperada, e estes estariam supostamente associados ao processo de senescência e antese do opérculo das flores, no momento de suas aberturas.

O *cluster* com maior representatividade observado na biblioteca de pistilos de *E. grandis* (*Cluster 002*, Tabela 4) apresentou similaridade a genes de *A. thaliana* codificadores de proteínas ricas em prolina. Em trabalhos de avaliação do perfil de expressão gênica em estigmas de *A. thaliana* durante o crescimento do tubo polínico conduzidos por Swanson *et al*. [80] e Tung *et al*. [81], o gene anotado como *At4g38770.1* correspondeu a uma extensina rica em prolinas. Diversas classes de extensinas têm sido descritas por outros autores em perfis transcricionais de estigma de flores e associados à interação pólen-pistilo [82-85].

Durante o processo de receptividade do estigma para a polinização, há um concerto de atividades exercidas pelas extensinas ricas em prolina e enzimas como pectinesterases e peroxidases. Estas últimas têm seu papel associado ao afrouxamento ou ao enrijecimento da parede celular. As peroxidases podem aumentar a extensibilidade da parede pela geração de radicais hidroxila oriundos da degradação da parede celular [86] ou podem aumentar a rigidez ao mediar uma rápida ligação às extensinas ricas em prolinas, como ocorre em danos físicos causados à planta ou durante a infecção por patógenos [81, 87]. A presença de peroxidases na superfície do estigma tem sido relacionada como um indicador de receptividade do estigma para a polinização de várias espécies, implicando estas enzimas no afrouxamento da parede durante a maturação do estigma [81]. Segundo estes mesmos autores, a expressão significativa de extensinas ricas em prolina em estigmas de *A. thaliana*, como encontrado em pistilos de *E. grandis*, sugere que a atividade da peroxidase papilar do estigma pode atuar no reforço da parede das células epidérmicas durante o crescimento do tubo polínico.

Com grande representatividade na biblioteca de anteras/estames, o *Cluster 003* apresentou mais alta homologia ao transportador de fosfato AtPT2 de *A. thaliana*, descrito por Mugge *et al*. [88] como único transportador de fosfato envolvido na

abscisão de órgãos florais neste vegetal. Este transportador pode promover a remobilização do fosfato inorgânico (Pi) nos órgãos florais em senescência para posterior transferência a flores recém-formadas, e a eliminação de qualquer Pi que eventualmente extravaze dos feixes vasculares expostos na zona de abscisão. Chapin & Jones [89] relacionaram a expressão de um transportador de fosfato, regulado pelo hormônio etileno, com o processo de senescência da corola de petúnia. Similarmente, a presença de um transportador de fosfato na biblioteca de anteras/estames de *E. grandis* pode estar relacionado com a queda dos estames após a fertilização da flor.

O *Cluster 004* mostrou-se similar ao transportador de inositol ATINT2 de *A. thaliana*, uma proteína integral de membrana envolvida no transporte de H^+/inositol através da membrana fosfolipídica da célula. Schneider *et al.* [90] observaram a expressão deste gene em anteras, tecido vascular e, em menor escala, no mesófilo. Nas anteras, este transportador pode estar envolvido na liberação de inositóis para o pólen em desenvolvimento, pois o mioinositol fosforilado fitato um dos principais constituintes dos grãos de pólen. Nos tecidos vasculares, sua atividade pode estar relacionada ao fornecimento de inositol para o galactinol e, portanto, para a biossíntese de rafinose. No mesófilo, sua função pode estar relacionada à recuperação de moléculas de inositol que se difundem para fora da célula. A presença de um transportador de inositol nas bibliotecas de anteras/estames e carpelos/receptáculos florais de *E. grandis* pode estar relacionado a liberação de inositóis para o pólen e a recuperação de moléculas difundidas no meio extracelular.

O processo de abscisão é considerado crítico no ciclo de vida de uma planta, atuando sobre uma série de órgãos, incluindo folhas, flores e frutos, como consequência de eventos altamente coordenados [91]. O início do processo ocorre em locais pré-determinados, denominados zonas de abscisão e, durante a separação do órgão, observa-se a dissolução da parede celular no ponto de desprendimento [92, 93]. A degradação da parede celular durante a abscisão de folhas, flores e frutos está associada ao aumento da atividade de várias enzimas hidrolíticas, incluindo β-1,4 glicanases [94-96] e poligalacturonases [94, 97, 98]. Genes codificadores destas duas classes de proteínas foram expressos em bibliotecas de flores de *G. hybrida*, *P. x hybrida* e flores masculinas de *P. fascicularis*. Em *E. grandis*, o *cluster 005* é constituído

por transcritos potencialmente codificadores de poligalacturonases, e estes foram encontrados nas bibliotecas de anteras/estames e carpelos/receptáculo floral. À semelhança dos demais trabalhos, estas poligalacturonases devem atuar na queda dos estames e do pistilo, no processo inicial da formação do fruto.

O *cluster 008* apresentou representatividade significativa junto à biblioteca de estames/anteras, com similaridade a um membro da superfamília RAS de proteínas associadas à secreção de *A. thaliana*. A abreviatura RAS deriva, do inglês, de *RAt Sarcoma*, e usualmente designa pequenas proteínas de hidrólise de DTP, ou GTPases. A formação de vesículas de transporte na célula depende de um mecanismo de seleção de cargas e concentrações junto às organelas celulares, especialmente retículo endoplasmático e complexo de Golgi, necessárias para definir o que pertence à vesícula. Associado a este evento, alguns sinais devem ser enviados a partir do lúmen para o citoplasma, onde se forma um revestimento das membranas (do inglês, *coat*) para indicar o local de formação da vesícula. Para tanto, um grande número de proteínas, conhecidas como proteínas de revestimento (do inglês, *coat proteins*), distintas entre si, são responsáveis pela formação das vesículas em um determinado ponto das membranas das organelas. As famílias de GTPases RAB e ARF são os principais componentes da maquinaria de tráfego de vesículas. A família RAB é a maior família de pequenas proteínas ligantes a GTP e que desempenham um importante papel na especificação da identidade e do tráfego de membrana da vesícula, ao longo de ambas as rotas exocítica e endocítica [99, 100]. A família ARF foi inicialmente identificada como fatores de ribosilação por ADP, e são reguladores da brotação de vesículas em diferentes etapas do tráfego de membrana. Os membros da família ARF estão subdivididos em subfamílias denominadas RAS, ARF e ARL. A subfamília RAS é necessária para o transporte das proteínas de revestimento do complexo II do retículo endoplasmático para o aparelho de Golgi. Também atuam na regulação das enzimas que metabolizam lipídios. No estudo de proteômica de micrósporos maduros de *A. thaliana*, proteínas da subfamília RAS foram identificadas e associadas à regulação da polaridade do crescimento do tubo polínico [101, 102]. Estes fatos permitem sugerir que a presença de proteínas RAS em anteras/estames de *E. grandis* esteja relacionada, pelo menos em parte, ao crescimento do tubo polínico.

O *cluster 006* encontra-se entre os que possuem maior representatividadedas bibliotecas de anteras/estames e carpelos/receptáculos florais. Conforme homologia encontrada frente ao TAIR e ao GenBank, o peptídeo codificado apresentou mais alta similaridade à enzima aldeído desidrogenase 2C4 de *A. thaliana* (ALDH2C4) e com a enzima aldeído desidrogenase de *R. communis*. Os aldeídos são intermediários em vários processos metabólicos fundamentais, incluindo a síntese de carboidratos, vitaminas, esteróides, aminoácidos, lipídios e fenilpropanóides. Suas moléculas variam de tamanho e nas caraterísticas de suas cadeias de alquila e são, geralmente, tóxicas aos sistemas biológicos devido à sua alta reatividade química. O acúmulo de aldeídos em órgãos vegetais é particularmente induzido em resposta a estresses ambientais como salinidade, desidratação, ressecamento, frio e choque térmico [103].

Os níveis de aldeídos no interior das células devem ser muito bem regulados e este papel cabe a enzimas aldeído desidrogenases, um grupo evolutivamente bem conservado de enzimas que catalisam a oxidação irreversível de uma ampla gama de moléculas endógenas aldeído-reativas para seus correspondentes ácidos carboxílicos, utilizando NAD^+ ou $NADP^+$ como cofatores. Em plantas, as enzimas aldeído desidrogenases estão envolvidas na tolerância a estresses abióticos, na restauração da esterilidade masculina, no desenvolvimento embrionário, na viabilidade das sementes e na maturação. A enzima ALDH2C4, localizada no citosol, tem como uma de suas funções fisiológicas a produção de ácido ferúlico e ácido sináptico durante a síntese de ligninas. A avaliação da expressão do gene codificador desta enzima realizada por Wei *et al*. [104] mostrou uma expressão constitutiva em raízes, folhas em roseta, caule, folhas caulinares, flores e síliquas, com diferentes níveis entre os tecidos. Diante deste perfil, é esperada a expressão deste gene nas bibliotecas de anteras/estames e carpelos/receptáculos florais de *E. grandis*, cujo papel está supostamente relacionado à homeostase celular.

Transcritos potencialmente codificadores da proteína calmodulina foram encontrados com uma representatividade significativa na biblioteca de carpelos/receptáculos florais de *E. grandis* (*cluster 009*). A calmodulina é uma proteína ubíqua em eucariotos, sendo a principal receptora intracelular de íons cálcio (Ca^{2+}). Ela

converte o sinal do mensageiro secundário Ca^{2+} ligando-se e alternando a atividade de proteínas-alvo gerando, assim, respostas fisiológicas a estímulos como a transdução de sinais, integrando seus efeitos aos de outras vias de sinalização [105-107] e, também, na manutenção do equilíbrio homeostático do Ca^{2+} para minimizar seus efeitos citotóxicos [105].

Em plantas, uma estreita relação existe entre o conteúdo e o estado celular de cálcio e o regulador de crescimento etileno. Por exemplo, o cálcio exerce crítica influência sobre a transcrição e a atividade da enzima 1-aminociclopropano-1-ácido carboxílico-oxidase (ACC), envolvida na síntese de etileno. Segundo trabalhos de Raz e Fluhr [108], o cálcio mostrou-se essencial para a ativação de proteínas relacionadas à patogênese como quitinases, sendo que a elevação dos níveis de cálcio são promovidos, pelo menos em parte, pelo etileno. Há muito se conhece os efeitos do etileno gasoso sobre o crescimento e o desenvolvimento vegetal, incluindo a indução do amadurecimento de frutos climatéricos, senescência, morte celular programada, abscisão de vários órgãos, promoção da germinação de sementes e a promoção ou a inibição da floração [109]. Assim, uma das mais prováveis funções dos genes codificadores de calmodulina em carpelos/receptáculos florais de *E. grandis* estaria relacionada à transição destas partes da flor em primórdios do fruto.

Além dos *clusters* já descritos, uma ampla gama de transcritos representados em menor proporção foram encontrados em flores de *E. grandis*, e estão potencialmente associados ao transporte intracelular, à montagem de proteínas, a reações de oxidação-redução, à resposta a estresses ou envolvidos em processos metabólicos (síntese) e de regulação da transcrição, conforme apresentados na Tabela 4. Por exemplo, associados a transporte intracelular estão os peptídeos potencialmente codificados pelos transcritos dos *clusters 004* (transportador de carboidratos), *067* e *069* (tubulinas), *070* (ATPase de membrana plasmática), e *071* (proteína 725 vesícula associada à membranas). Os *clusters 035* (proteína de choque térmico), *038* (superóxido dismutase), *042* (chaperonina), *046* (lipoxygenase), *053* (peroxidase), *075* (catalase), *044* (proteína de senescência), *090* (ubiquitina), *084* (GTP-binding family), *096* (endoquitinase acídica) estão possivelmente associados aos

processos celulares de montagem de proteínas, óxidação-redução e respostas a estresses. Vinculados à síntese e demais processos metabólicos estão os transcritos integrantes dos *clusters 010* (GDSL lipase), *082* (peroxisomal acil-CoA oxidase), *073* (malato oxirredutase), *005* (poligalacturonase), *024* (liases de carbono-enxofre), *030* (proteína fosfotirosina fosfatase), *031* (pectinacetiltransferase), *033* (NADH desidrogenase), *048* (xiloglicanaendotransglicosilase), *068* (aspartil protease), *045* (glutamina sintetase), *049* e *055* (proteína ribossomal 60S), e *036* (folilpoliglutamato sintase). Finalmente, diversos transcritos potencialmente codificadores de fatores de transcrição foram encontrados. Conforme anteriormente comentado, a relevância destas proteínas sobre a síntese de RNAs e, consequentemente, sobre a síntese de todas as demais proteínas, torna-as especiais, mesmo que os níveis de transcritos para as mesmas não tenham sido proporcionalmente os mais altos. Assim, alguns dos *clusters* potencialmente codificadores de fatores de transcrição serão descritos na próxima sessão.

Fatores de Transcrição Expressos em Flores de *E. grandis*

A biologia de plantas é comum em muitos aspectos ao de outros organismos. Entretanto, uma série de processos biológicos são exclusivos de plantas como, por exemplo, a fotossíntese, a assimilação de nitrogênio, o processo reprodutivo e as respostas a estímulos ambientais [110]. As flores são estruturas complexas, consistindo de vários órgãos como sépalas, pétalas, estames e carpelos. O desenvolvimento da flor inclui a formação e manutenção das inflorescências e meristemas florais, a especificação da identidade dos órgãos florais, iniciação dos primórdios florais e subseqüente organogênese [111, 112].

Vários estudos genéticos e moleculares em *Arabidopsis* e outras espécies de plantas identificaram genes que desempenham papéis chaves durante do desenvolvimento e manutenção da flor. Muitos destes genes codificam fatores de transcrição, proteínas quinases e proteínas dedo de zinco [113].

A regulação da transcrição é considerado um dos passos mais estudados na modulação da expressão gênica, uma vez que, de todos as etapas envolvidas, apenas o controle da transcrição assegura que nenhum intermediário supérfluo seja sintetizado [114]. Os fatores de transcrição são proteínas responsáveis pela seletividade na regulação genética e, muitas vezes, são expressos em tecidos ou estádios específicos do desenvolvimento, ou por caminhos dependentes de estímulos exógenos e/ou endógenos, durante todo o ciclo de vida da planta [115-117].

Das quatro bibliotecas de flores de *E. grandis* foram selecionados 21 grupos de transcritos potencialmente codificadores de fatores de transcrição. Estes grupos foram classificados em onze famílias, de acordo com as anotações obtidas junto aos banco de dados *Plant Transcription Factor Database* da Universidade de Potsdam, Alemanha [118] e da Universidade de Pequim, China) [119]. Do total de 21 fatores de transcrição, nove foram encontrados expressos nas bibliotecas de carpelos/receptáculos florais; oito expressos em estames/anteras; três em pistilos e um em sépalas/pétalas. Destes, três formaram clusters com dois transcritos cada (Tabela 4) e 18 foram representados uma única vez no transcriptoma (*singletons*, Tabelas 5).

Os fatores de transcrição encontrados em flores de *E. grandis* pertencem às famílias AP2, B3, zf-C3HC4, Dof, ERF, GATA, MADS, MYB, LIM, NAC e WRKY. A biblioteca de estames/anteras foi a que apresentou a maior diversidade de fatores de transcrição, com genes expressos representando as famílias B3, ERF, GATA, MADS, MYB, LIM e WRKY. Os fatores de transcrição MYB, LIM e WRKY foram encontrados exclusivamente nesta biblioteca. As famílias AP2, B3, zf-C3HC4, Dof, ERF e NAC foram expressas nas bibliotecas de carpelos/receptáculos florais, e os fatores de transcrição zf-C3HC4, Dof e NAC foram exclusivos desta biblioteca. Na biblioteca de pistilo foram expressos os fatores pertencentes às famílias AP2, B3 e MADS. A única sequência encontrada na biblioteca de sépalas/pétalas correspondente a um fator de transcrição pertencente à família GATA (Tabela 5). Novamente, os transcritos potencialmente codificadores de fatores de transcrição foram anotados segundo os bancos de dados TAIR e GenBank (NCBI). Entretanto, a descrição das prováveis funções de cada fator de

transcrição foi realizada tendo-se por base apenas as anotações do banco de dados de *A. thaliana* (TAIR), por ser considerado um banco mais robusto.

O *cluster 106* (Tabela 4) apresentou seqüência consenso codificadora de peptídeo similar ao fator de transcrição que contém o domínio AP2 (At4g34410.1). Estes fatores desempenham papel central na corregulação da rede conhecida por *Redox Relugated Gene 1* (RRG1), associada à manutenção da homeostase e da capacidade de respostas redox [120]. Nos estudos de Feng *et al.* [121], a transcrição do fator AP2 At4g34410.1 mostrou-se aumentada em 10,6 vezes após o tratamento com NaCl 19,2 vezes maior em condições de seca e 4,7 vezes maior quando a planta foi submetida à radiação UV. Entretanto, a expressão do mesmo foi ligeiramente diminuída sob o tratamento com ABA e insensível a frio e calor, indicando funções específicas dos circuitos de regulação redox na resposta ao estresse de *A. thaliana*. Este gene também foi associado à resposta vegetal à infecção por patógenos míldios de *A. thaliana* e sua expressão aumentou em resposta ao aumento da biomassa do patógeno [122]. O gene de At4g34410.1 respondeu ao estresse salino no processo de senescência das folhas desta espécie [123]. Em flores de *E. grandis*, dois transcritos para este gene foram encontrados em carpelos/receptáculos florais, provavelmente atuando na resposta redox da flor diante do processo de senescência que este órgão sofre, dando início à formação do fruto.

Pertencente a mesma família de fatores de transcrição AP2, o *cluster 111* (Tabela 4) apresentou similaridade com a sequencia de peptídeos da proteína At3g19184.1. Esta proteína de ligação ao DNA foi designada como RAV1 (do inglês, RAV: for Related to ABI3/VP1) e contém um domínio AP2 DNA-binding na região N-terminal e um domínio B3 altamente conservado, que caracteriza os fatores de transcrição VP1/ABI3, na região C-terminal da proteína [124, 125]. Os domínios de ligação ao DNA desta proteína atuam cooperativamente para alcançar uma alta afinidade de ligação. Estes fatores de transcrição são os primeiros exemplos de proteínas contendo dois domínios distintos de ligação ao DNA e encontrados exclusivamente em plantas. Entretanto, ainda não se conhece suas funções fisiológicas e tão pouco os genes-alvos destas proteínas, mas acredita-se que sua função pode

estar relacionada a algum processo fisiológico que evoluiu com as plantas superiores [125, 126].

Como ocorre com muitos processos de desenvolvimento e resposta a estímulos ambientais nas plantas, a expressão gênica desempenha um papel importante na transição de um órgão funcional à senescência. Os fatores de transcrição agem como um interruptor para fazer com que a expressão de genes por ligação a elementos específicos *cis* dos promotores do gene alvo, resultando na ativação e/ou repressão destes genes. Segundo estudos de Guo *et al*. [127] identificaram 20 famílias de fatores de transcrição expressas durante o processo de senescência de folhas de *A. thaliana*. Alguns fatores de transcrição representantes destas famílias foram encontrados nas bibliotecas de flores, como por exemplo, NAC, WRKY, MYB e AP2.

Os fatores de transcrição NAC (*clusters 037* e *047*, Tabela 4) são específicos de plantas e estão envolvidos em diversos processos da planta. Algumas análises de expressão gênica desta família de fatores de transcrição mostram a participação destes genes no desenvolvimento da flor e na reprodução [125-127], bem como em resposta a hormônios [130, 131] e estresses bióticos [132] e abióticos [130, 133, 134]. Os genes NAC também foram implicados na da planta à luz, no processo de morte celular programada e na senescência. Portanto, a presença de dois *clusters* similares a genes pertencentes à família NAC na biblioteca de carpelos/receptáculos florais de *E. grandis* pode estar relacionado ao processo de senescência e transição da flor para a frutificação.

Tabela 5 - **Anotação *in silico* por BlastX frente aos bancos TAIR e GenBank (NCBI) de fatores de transcrição, representados por *singletons*, obtidos a partir de ESTs válidas das bibliotecas de flores de *E. grandis*.** As linhas destacadas em verde representam os genes cuja expressão foi avaliada por RT-qPCR.

Família	*Singleton* ID	TAIR			NCBI			Parte da Flor
		ID	Anotação (Melhor *hit* no BlastX)	e-value	ID	Anotação (Melhor *hit* no BlastX)	e-value	
AP2/ERF	*EugrTF01*	AT5G38560.1	Protein kinase superfamily protein	1e-27	ref|XP_002305965.1|	predicted protein [*P. trichocarpa*]	3e-45	P
B3	*EugrTF02*	AT1G15750.4	WSIP1, TPL Transducin family protein	2e-55	ref|NP_563981.1|	PL (TOPLESS); protein binding/homodimerization transcription repressor [*A. thaliana*]	2e-53	C/RF
	EugrTF03	AT3G48050.1	Zinc finger C-x8-C-x5-C-x3-H type family protein	3e-40	emb|CAN60153.1|	hypothetical protein ITISV_021504 [*V. vinifera*]	2e-84	A/E
	EugrTF04	AT2G41900.1	CCCH-type zinc finger protein with ARM repeat domain	2e-61	gb|ABN08215.1|	Zinc finger, CCCH-type [*M. truncatula*]	8e-68	P
zf-C3HC4	*EugrTF05*	AT2G46160.1	RING/U-box superfamily protein	1e-27	gb|ABD28581.1|	Zinc finger, RING-type [*M. truncatula*]	1e-25	C/RF

Família	*Singleton* ID	TAIR			NCBI			Parte da Flor
		ID	Anotação (Melhor *hit* no BlastX)	e-value	ID	Anotação (Melhor *hit* no BlastX)	e-value	
Dof	*EugrTF06*	AT4G3800.1	DOF4.7 DNA binding with one finger 4.7	3e-23	ref|XP_002509452.1|	zinc finger protein, putative [*R. communis*]	9e-45	C/RF
	EugrTF07	AT5G39660.2	CDF2 (CYCLING DOF FACTOR 2); DNA binding / transcription factor	2e-79	XP_002283706.1	PREDICTED: hypothetical protein [*V. vinifera*]	5e-17	C/RF
ERF	*EugrTF08*	AT5G61890.1	Integrase-type DNA-binding superfamily protein	1e-30	ref|XP_002275870.1|	hypothetical protein [*V. vinifera*]	3e-10	C/RF
	EugrTF09	AT2G44840.1	ethylene-responsive element binding factor 13	2e-27	ref|XP_002327050.1|	AP2/ERF domain-containing transcription factor [*P. trichocarpa*]	7e-33	A/E
GATA	*EugrTF010*	AT3G54250.1	GHMP kinase family protein	6e-90	ref|XP_002521172.1|	diphosphomevalonate decarboxylase, putative [*R. communis*]	4e-88	S/P
MADS	*EugrTF011*	AT5G23540.2	Mov34/MPN/PAD-1 family protein	2e-85	ref|XP_002302444.1|	predicted protein [*P. trichocarpa*]	4e-85	A/E
	EugrTF012	AT1G24260.3	SEP3| K-box region and MADS-box transcription factor	1e-04	gb|AAN15182.1|	MADS box protein GHMADS-1 [*G. hirsutum*]	2e-10	P
	EugrTF013	AT2G45660.1	AGL20, SOC1, ATSOC1 | AGAMOUS-like 20	8e-31	gb|ABX90014.1|	SOC1-like protein 1 [*Sinningia speciosa*]	9e-32	C/RF
MYB	*EugrTF014*	AT4G32730.1	MYB3R1 Homeodomain-like protein	4e-11	ref|XP_002516567.1|	myb, putative [*R. communis*]	1e-12	A/E
LIM	*EugrTF015*	AT1G10200.1	WLIM1 GATA type zinc finger transcription factor	2e-94	gb|ACN97188.1|	LIM transcription factor [*P. trichocarpa*]	2e-96	A/E

Família	*Singleton* ID	TAIR			NCBI			Parte da Flor
		ID	Anotação (Melhor *hit* no BlastX)	e-value	ID	Anotação (Melhor *hit* no BlastX)	e-value	
WRKY	*EugrTF016*	AT2G04880.2	ZAP1, ATWRKY1, WRKY1 zinc-dependent activator protein-1	2e-28	gb|ACV92004.1|	WRKY transcription factor 2 [*P. tomentosa* x *P. bolleana*]	5e-36	A/E

* Parte da Flor: S/P – pétalas/sépalas; A/E – anteras/estames; P – pistilos; C/RF – carpelos/receptáculos florais.

Os fatores de transcrição da família WRKY, representado pelo *singleton EugrTF016* (Tabela 5), expresso na biblioteca de anteras/estames de *E. grandis*, são proteínas que contém um ou dois domínios altamente conservados WRKY, caracterizado pelo heptapeptídeo WRKYGQK e uma estrutura dedo de zinco distinto dos demais fatores de transcrição que possuem esta estrutura. Em plantas, muitas proteínas WRKY estão envolvidas na defesa contra a infecção por bactérias patogências [135-140], fungos [139], vírus [139, 141, 142] e oomicetos [139, 143, 144]. Além disso, estes fatores de transcrição estão implicados na respostas ao estresse abiótico e injúria por ferimento [145], estresses pela combinação de seca e calor [146] e frio [147, 148].

Alguns membros da família desempenham papéis importantes na regulação da morfogênese dos tricomas [149] e embriões [150], além de participarem de processos importantes na fisiologia da planta, como na senescência [140, 151-153], dormência [147], crescimento [140] e vias metabólicas [149, 154-156]. Diante deste panorama, acredita-se que a expressão deste fator de transcrição na biblioteca de anteras/estames de *E. grandis* esteja envolvido no processo metabólico da flor.

O *singleton EugrTF015* (Tabela 5), presente na bibliteca de anteras/estames de *E. grandis*, apresentou sequencia consenso codificadora de peptídeo similar ao fator de transcrição que contém o domínio LIM, descrito para diversas espécies de plantas como tabaco [157], girassol [158, 159] e *Arabidopsis* [160]. Em tabaco, e possivelmente em todas as plantas superiores, existem duas isoformas: uma específica para os grãos de pólen (PLIM) e uma expressa em todos os órgãos vegetativos e reprodutivos (WLIM) [160]. A proteina WLIM se mostrou associada a plastídios nas células epidérmicas das folhas e do caule, bem como na região perinuclear das células meristemática da ponta da raiz [158].

O domínio LIM serve de local para interação entre proteína-proteína formando homodímeros [161], com fatores de transcrição tipo *helix-loop-helix* [162, 163], ou com proteínas quinases [164]. Muitas proteínas contendo o domínio LIM estão implicadas na regulação da transcrição da diferenciação celular e regulação do crescimento. Algumas destas proteínas estão associadas ao citoesqueleto, enquanto

outras estão envolvidas na translocação cromossômica [110]. Portanto, é provável que o fator de transcrição WLIM expresso na biblioteca de anteras/estames esteja associado ao citoesqueleto celular.

Nas bibliotecas de anteras/estames e pistilo de flores *E. grandis* foram encontratos três sequencias prováveis codificadoras de membros da família MADS-box de fatores de transcrição, que inclui a maioria de genes necessários para o modelo ABC de determinação da identidade de órgãos florais em desenvolvimento [63]. Os genes desta família controlam diversos processos de desenvolvimento, incluindo raízes, frutos e flores, e estão envolvidos com a homeostase do sistema floral [63, 165].

Dois *singletons* distintos (*EugrTF06* e *EugrTF07*) codificadores de potenciais proteínas Dof foram encontrados entre os transcritos de flores de *E. grandis* (Tabela 5). As proteínas Dof constituem uma família de fatores de transcrição exclusiva de plantas, que contém uma classe particular de domínio de ligação ao DNA do tipo dedo-de-zinco. A abreviatura Dof deriva, do inglês, de *DNA one finger*, exatamente para designar este peculiar domínio [166]. Análises bioquímicas, moleculares e genéticas revelaram que proteínas Dof funcionam como ativadoras ou repressoras transcricionais em diversos processos biológicos em vegetais como, por exemplo, na resposta a estresses oxidativos [167, 168], em mecanismos de defesa [169], na regulação de genes associados ao metabolismo de carboidratos em resposta à luz [170-172], em resposta à sinalização por fitocromos [173], em resposta a fitormônios incluindo auxinas [174, 175] e giberelinas [176, 177], germinação de sementes [178, 179], síntese de proteínas de reserva no endosperma em desenvolvimento [180], e na regulação específica de estômatos [181]. Em virtude de dois distintos transcritos potencialmente codificadores de fatores Dof terem sido encontrados em flores de *E. grandis*, da variedade de genes regulados pelos mesmos em plantas e pela relativa inexistência de informações com respeito a genes codificadores destes fatores em espécies arbóreas, além do estudo destes fatores de transcrição ser uma linha de pesquisa do Laboratório de Biologia Molecular Vegetal, um capítulo específico foi redigido junto a esta Tese, com dedicação aos genes *Dof* de *E. grandis* (Capítulo 2).

Avaliação da Expressão Espacial de Genes Expressos nas Bibliotecas de Flores de *E. grandis*

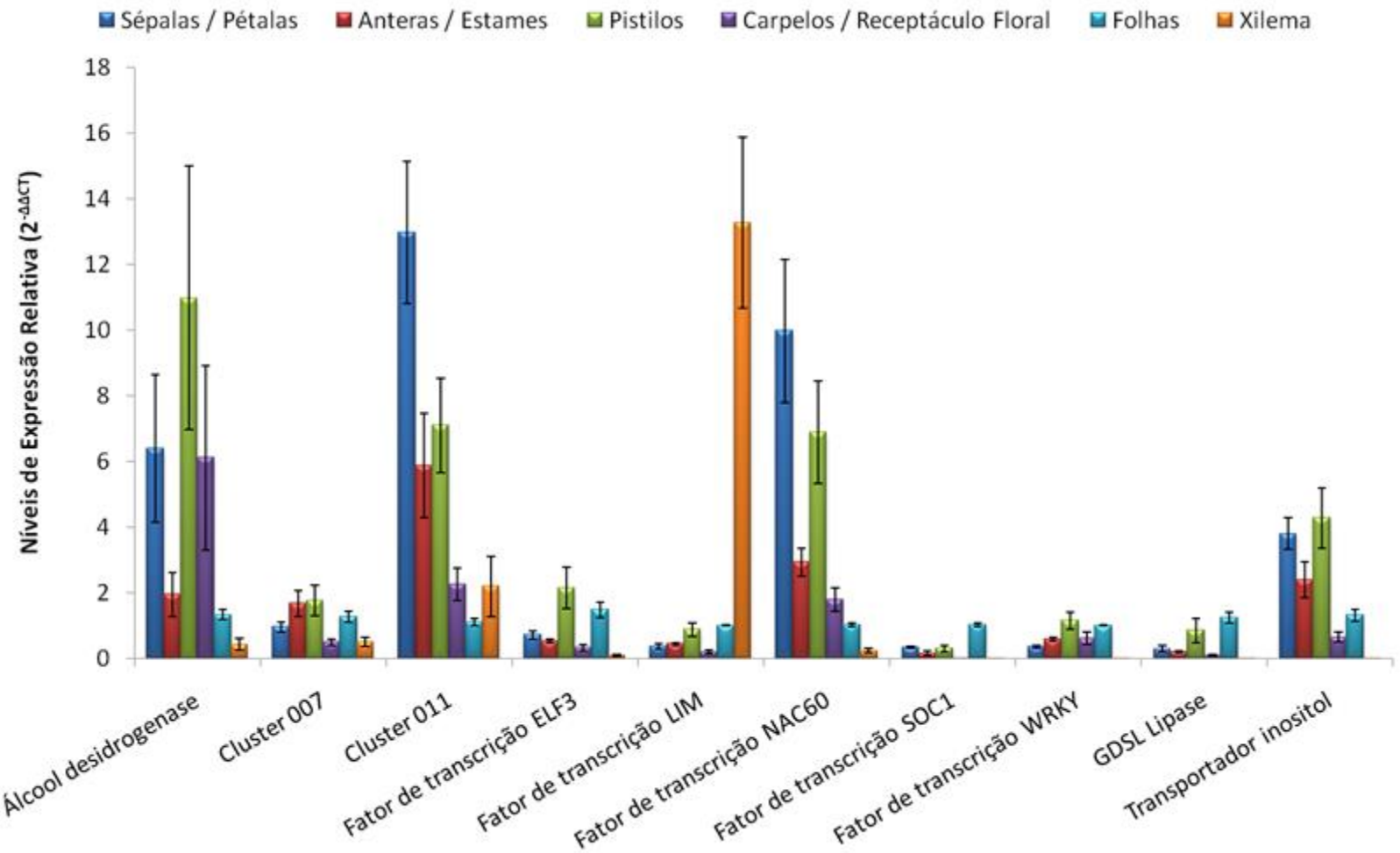
A partir das bibliotecas de partes da flor de *E. grandis* supracitadas foram selecionados dez genes envolvidos em diversos processos celulares e fisiológicos distintos para análise de expressão. Adotou-se como critério de seleção, genes preferencialmente expressos nas bibliotecas de anteras/estames e/ou carpelos/receptáculos florais, a maioria repesentados por *clusters*. Foram também selecionados genes não caracterizados que se apresentaram que possuem maior representatividadenas análises de agrupamentos de sequencias. Os oligonucleitídeos correspondentes a estes genes tiveram sua especificidade de amplificação avaliada por meio da curva de dissociação. Os níveis de expressão dos genes avaliados foram comparados aos valores de expressão obtidos em folhas, cujo nível de expressão está representado no gráfico com o valor próximo a 1 (Figura 6).

A análise de expressão dos dez genes selecionados foi realizada em flores particionadas, folhas e xilema de *E. grandis*. Deste total, observou-se que 9 genes foram diferencialmente expressos em tecidos de sépalas/pétalas, 8 em anteras/estames, 5 em pistilos, 8 em carpelos/receptáculos florais e 8 no tecido xilemático. Dentre os genes analisados, os fatores de transcrição NAC60 e SOC1, e o trasportador de inositol foram os que apresentaram níveis de expressão aumentados ou diminuídos em todas as partes da flor (Figura 6).

O gene com similaridade à proteína álcool desidrogenase de cadeia curta de *A. thaliana*, apresentou altos níveis de expressão em pistilos, sépalas/pétalas e carpelos/receptáculos florais, embora sua expressão tenha sido observada na biblioteca de anteras/estames e carpelos/receptáculos florais. No tecido xilemático, sua expressão se mostrou reduzida a metade em relação a folhas. Em estudo sobre o processo de senescencia foliar, Gan e Amasino [182] citam o gene correspondente a proteína álcool desidrogenase de cadeia curta como um dos genes cuja expressão

aumentada durante a senescência, portanto, era esperado que o perfil de expressão deste gene estivesse aumentado nos tecidos citados.

Figura 6 – Níveis de expressão relativa de 10 genes expressos nas bibliotecas de flores de *E.* grandis, por RT-qPCR. Os perfis de expressão foram obtidos para RNAs de flores (sépalas/pétalas, anteras/estames, pistilos e carpelos/receptáculos florais), folhas e xilema de *E.* grandis. Os valores de expressão foram normalizados em relação aos valores de folhas. *Histona H2B*, *At2g28390* e *G3PDH* foram utilizados como genes referência para a normalização da expressão. As barras representam o erro padrão.



Dentre os representantes dos *clusters* que possuem maior representatividadedas bibliotecas de anteras/estames e carpelos/receptáculos florais, o perfil de expressão apresentado pelo *cluster 007* se mostrou similar ao encontrado em folhas para os tecidos de anteras/estames e pistilos, e reduzido para carpelos/receptáculos florais e xilema, enquanto que o perfil de expressão do *cluster 011* apresentou um elevado aumento de expressão para os tecidos de sépalas/pétalas, anteras/estames e pistilos.

O perfil de expressão dos genes similares as proteínas *ELF3* e GDSL lipase, identificados nas bibliotecas de pistilo e carpelos/receptáculos florais, apresentaram semelhança quanto aos tecidos analisados. Ambos apresentaram redução nos níveis de transcrição para os tecidos de sépalas/pétalas, anteras/estames, carpelos/receptáculos florais e xilema, enquanto que os valores encontrados para pistilos foram similares aos de folhas. Pertencente a uma importante família de lipases, as GDSL lipases atuam ativamente na hidrólise e síntese de lipídios ou ésteres. Fisiologicamente, as GDSL lipases estão envolvidas na regulação do crescimento e desenvolvimento das plantas, resistência à doenças e na germinação de sementes [183]. Com um nível transcricional em pistilo semelhante ao encontrado em folha, e identificado na biblioteca de carpelos/receptáculos florais, este gene pode estar associado a regulação do crecimento e desenvolvimento da flor. O papel da proteína Early Flowering 3 (ELF3) na planta é controlar o florescimento através da regulação do gene Constans (CO) [184]. De acordo com o esperado, os níveis de transcritos deste gene nos tecidos florais e xilema foram baixos, pois este gene tem sua expressão máxima antes da formação da flor.

Os valores de expressão observados em xilema para o gene similar ao fator de transcrição LIM foram os mais elevados dentre todos os genes e tecidos avaliados. Entretanto, os valores de expressão para os tecidos de sépalas/pétalas, anteras/estames e carpelos/receptáculos florais foram inferiores aos encontrados em folhas. Como já descrito, as proteínas que contém o domínio LIM atuam na regulação da transcrição e do crescimento, e estão associadas ao citoesqueleto celular, e portanto, sua elevada expressão em xilema de *E.grandis* pode estar relacionado ao crescimento da planta.

Como já mencionado, os genes similares aos fatores de transcrição NAC60 e SOC1 e o transportador de inositol foram os únicos cujo perfil de expressão se apresentou diferenciados em todos os tecidos de flor. O fator de transcrição NAC60 mostrou aumento na expressão nos tecidos de flor e redução no tecido xilemático; o transportador de inositol mostrou aumento para os tecidos de sépalas/pétalas, anteras/estames e pistilos, e redução em xilema; enquanto que o fator de transcrição SOC1 apresentou redução em todos os tecidos analisados quando comparado com folhas. As proteínas codificadas pelos genes pertencentes à família de fatores de transcrição NAC, a exemplo do que foi descrito acima, estão associados ao processo de senescencia da flor, e portanto, é esperado que sua expressão tenha maiores níveis do que o encontrado em folhas e xilema, devido ao processo de transição da flor para a frutificação. O gene similar a um transportador inositol tem como uma de suas prováveis funções nos tecidos de sépalas/pétalas, anteras/estames e pistilos, a liberação de inositóis para biossíntese e a recuperação de moléculas difundidas no meio extracelular. O fator de transcrição SOC1 codifica uma proteína MADS-box e é bem conservada entre as espécies de angiospermas [185-189]. Estudos recentes têm demonstrado que esta proteína é multifuncional, que regula não só o tempo de florescimento, mas também a padronização da flor e a determinação do meristema floral [190-193]. Portanto SOC1 pode desempenhar um papel como regulador geral da organogênese no desenvolvimento da planta, e por isto, era esperado que os maiores níveis de expressão deste gene fossem encontrados em folhas.

O gene correspondente ao fator de transcrição WRKY apresentou um nível de expressão semelhante ao encontrado em folhas nos tecidos de pistilos e carpelos/receptáculos florais e reduzidos em sépalas/pétalas, anteras/estames e xilema. Estas proteínas estão envolvidas em muitos processos na planta, dentre eles, a defesa contra a infecção por bactérias patogênicas [135-140], fungos [139], vírus [139, 141, 142] e oomicetos [139, 143, 144]. Além disso, estes fatores de transcrição respondem ao estresse abiótico, injúrias [145], estresses pela combinação de seca e calor [146] e frio [147, 148]. Portanto os baixos níveis de expressão deste fator de transcrição nos tecidos florais e xilema estão dentro do esperado, uma vez que a flor não está, aparentemente, sob nenhum estresse para desencadear sua expressão.

CONCLUSÕES

- um total de 3.190 ESTs foram geradas ao todo, totalizando 1.641 ESTs consideradas válidas, distribuídas em 120 *clusters* e 2.830 *singletons*;
- as sequências foram assim distribuídas entre as bibliotecas: 69 unigenes estão presentes na biblioteca de sépalas/pétalas, 136 em anteras/estames, 68 em pistilos e 130 em carpelos/receptáculo floral;
- o maior número de ESTs identificadas referem-se a genes de expressão constitutiva ou a genes potencialmente relacionados aos sinais fisiológicos de senescência da flor;
- os fatores de transcrição encontrados em flores de *E. grandis* pertencem às famílias AP2, B3, zf-C3HC4, Dof, ERF, GATA, MADS, MYB, LIM, NAC e WRKY;
- dos genes selecionados para avaliação da expressão por RT-qPCR, observou-se que 9 genes foram diferencialmente expressos em tecidos de sépalas/pétalas, 8 em anteras/estames, 5 em pistilos, 8 em carpelos/receptáculos florais e 8 no tecido xilemático;
- dentre os genes analisados, os fatores de transcrição NAC60 e SOC1, e o trasportador de inositol foram os que apresentaram níveis de expressão aumentados ou diminuídos em todas as partes da flor.

METODOLOGIA

Material biológico, purificação de RNA e construção das bibliotecas de cDNA de flores de *E. grandis*

Flores de *E. grandis* com quatro anos de idade, em início de antese, foram divididas em quarto partes, assim denominadas: pétalas/sépalas, anteras/estames, pistilos e carpelos/receptáculos florais.

O RNA total foi extraído das amostras utilizando o *kit Pure Link Plant RNA Reagent* (Invitrogen), de acordo com as instruções do fabricante, totalizando 5 µg de RNA total por amostra. Para a construção das bibliotecas de cDNA, foi utilizado o *kit SMART cDNA Library Construction* (Clonthec Laboratories).

A extração de DNA plasmidial foi realizada em microplacas de 96 poços utilizando o método de *boiling prep*, baseado na lise das paredes bacterianas por um detergente não iônico, lisozima e calor (banho-maria). As amostras de DNA foram amplificadas (30 a 45 ng) com 1 µL de *primer T7 forward* (5'-TGTAATACGACTCACTATAGGGCGA-3', Invitrogen) ou *M13-40 forward* (5'-GTTTTCCCAGTCACGACGTTGTA-3', Invitrogen) a concentração de 10 µM e 2 µL do *kit BigDye Terminator v3.1 Cycle Sequencing RR-100* (Applied Biosystems), em um volume final de reação de 10 µL. As reações foram realizadas no termociclador *GeneAmp 9700 PCR System termocycler* (Applied Biosystems), nas seguintes condições: 96°C por 3 minutos, seguido por 25 ciclos de 96°C por 10 segundos, 55°C por 5 segundos e 60°C por 4 min. Após a marcação, as amostras foram purificadas por precipitação com isopropanol, seguido por uma lavagem com álcool 70%. Os produtos da precipitação foram ressuspensos em 10 µL de formamida e denaturado a 95°C por 5 minutos, incubado em gelo e eletroinjetado no sequenciador automático. As amostras de DNA marcadas foram resolvidas no sequenciador automatic *ABI-PRISM 3100 Genetic Analyzer* com capilares de 50 cm, preenchidos com o polímero *POP6 polymer* (Applied Biosystems).

Os dados gerados no sequenciamento foram extraídos pelo programa *Data Collection software v1.0.1* (Applied Biosystems) e processados com o programa *CodonCode Aligner 3.0.3 for Windows* (CodonCode Corporation). As ESTs selecionadas para este estudo foram aquelas que obtiveram um tamanho mínimo de 150pb e uma qualidade de intensidade de sinal captada pelo software igual ou superior a 30 (considerada de boa qualidade).

Análises de bioinformática das bibliotecas de cDNA

O agrupamento das sequencias de cDNA das bibliotecas em *contigs* e *singletons* foi realizado também pelo programa *Codon Code*. Para estimar a similaridade dos *clusters* e *singletons* gerados, as sequencias foram submetidas a busca por similaridade com proteínas, pelo uso da ferramenta BlastX, a banco de dados públicos como NCBI, TAIR, Vitis and JGI Poplar. A anotação funtional dos *clusters*/*singletons* foi realizada pelo programa Blast2GO.

Análise da expressão por RT-qPCR

RNA total de flores, folhas e xilema de plantas de *E. grandis* com três anos de idade foram extraídos utilizando o *kit Pure Link™ Plant RNA Reagent* (Invitrogen), de acordo com instruções do fabricante. Para a síntese de cDNA, foi utlizado o *kit SuperScript III reverse transcriptase* (Invitrogen). As reações foram compostas por 10 µL cDNA diluído (1:50) a partir da reação de RT (RNA total a 1 µl), 2 µL *Platinum Taq* tampão para PCR 10X (Invitrogen), 1.2 µL $MgCl_2$ (50 mM), 0.4 µL dNTPs (5mM), 0.4 µL *primers* (10 µM), 4.95 µL de água, 1 µL *SYBR-Green* (Molecular Probes), 0.05 µL *Platinum Taq* DNA Polimerase (5 U/µL) (Invitrogen). A temperatura de anelamento utilizada foi de 60°C para todos os *primers,* projetados pela ferramenta *GenScript Corporation's Real-Time PCR Primer Design*. A reação de transcrição reversa quantitativa (*RT-qPCR*) foi realizada no equipamento *Applied Biosystem 7500 Real-time Cycler*. Os genes referência Histona, At2g28390 e G3PDH foram usados como controles internos para normalização os níveis de mRNA presente em cada amostra utilizando método $2^{-\Delta\Delta Ct}$ descrito por *Livak e Schmittgen* [194]. As amostras foram coletadas de 4

indivíduos de *E. grandis* diferentes e o delineamento do experimento foi feito com quatro replicadas experimentais.

REFERÊNCIAS BIBLIOGRÁFICAS

1. Pereira JCD, Sturion JA, Higa AR, Higa RCV, Shimizu JY: *Características da madeira de algumas espécies de eucalipto plantadas no Brasil*. Colombo PR: Embrapa; 2000:3-114.

4. Eldridge K, Davidson J, Harwood C, Van Wyk G: *Eucalypt domestication and breeding*. Oxford Univ Pr on Demand; 1994:288.

8. Siqueira L: **Variabilidade genética de uma população selecionada e teste de paternidade de uma progênie de polinização aberta de *Eucalyptus grandis* Hill ex Maiden**. *Agricultural and Forest Entomology* 2009, **74**:40.

10. Sousa VA, Higa RCV: **Fenologia de Florescimento e Frutificação de *Eucalyptus dunnii***. 1991:9-19.

33. Lorenzi H: *Árvores exóticas no Brasil: madeireiras, ornamentais e aromáticas*. Instituto Plantarum de Estudos da Flora; 2003.

34. Richards A: *Plant breeding systems*. Garland Science; 1986:529.

35. Boland DJ, Brooker MIH, Chippendale GM, et al. *Forest trees of Australia*. Fifth edit. Melbourne: Csiro Publishing; 1984:687.

36. Paludzyszyn Filho E, Pacheco AR, Dittmar H, Cordeiro CA: *Estratégias para o Melhoramento de Eucaliptos Trocipais na Embrapa*. Colombo - PR: 2004, **Documentos**.

37. Aguiar IB de, Kageyama PY: **Desenvolvimento floral de *Eucalyptus grandis* Hill Ex Maiden em Mogi Guaçu - SP**. *IPEF* 1987, **37**:5-11.

38. Williams J, Woinarski J: *Eucalypt ecology individuals to eosystems*. Cambridge University Press; 1997, **11**:413.

39. Assis FT: **Melhoramento genético do eucalipto**. *Informe Agropecuário* 1986, **12**:36-46.

40. Hodgson LM: **Some aspects o flowering and reproductive behaviour in *Eucalyptus grandis* (Hill) Maiden at J.D.M. Keet Forest Reearch station**. *South African Forestry Journal* 1976, **98**:32-43.

41. Schumacher MV, Calil FN, Vogel HLMV: *Silvicultura Aplicada*. Santa Maria - RS: 2005:120.

42. Fobel M, Lynch DV, Thompson JE: **Membrane Deterioration in Senescing Carnation Flowers**. *Photosynthetica* 1987:204-211.

43. Panavas T, Rubinstein B: **Oxidative events during programmed cell death of daylily (*Hemerocallis hybrid*) petals**. *Plant Science* 1998, **133**:125 - 138.

44. Fukuchi-mizutani M, Ishiguro K, Nakayama T: **Molecular and functional characterization of a rose lipoxygenase cDNA related to flower senescence**. *Plant Cell* 2000, **160**:129 - 137.

45. Alba R, Fei Z, Payton P, et al. **ESTs, cDNA microarrays, and gene expression profiling: tools for dissecting plant physiology and development.** *The Plant journal : for cell and molecular biology* 2004, **39**:697-714.

46. Meyers BC, Galbraith DW, Nelson T, Agrawal V: **Methods for transcriptional profiling in plants. Be fruitful and replicate**. *Plant Physiology* 2004, **135**:637.

47. Adams MD, Kelley JM, Gocayne JD, et al. **Complementary DNA sequencing: expressed sequence tags and human genome project** . *Science* 1991, **252** :1651-1656.

48. Liang P, Pardee AB: **Differential Display**. *Differential Display Methods and Protocols* 1997, **85**:3-11.

49. Velculescu VE, Zhang L, Vogelstein B, Kinzler KW: **Serial analysis of gene expression.** *Science (New York, N.Y.)* 1995, **270**:484-7.

50. Bachem CW, Hoeven RS van der, Bruijn SM de, et al. **Visualization of differential gene expression using a novel method of RNA fingerprinting based on AFLP: analysis of gene expression during potato tuber development.** *The Plant journal : for cell and molecular biology* 1996, **9**:745-53.

51. others: **AFLP: a new technique for DNA fingerprinting**. *Nucleic acids research* 1995, **23**:4407.

52. Diatchenko L, Laut Y-fai C, Campbellt AP, et al. **Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes**. *Proc. Natl. Acad. Sci.* 1996, **93**:6025-6030.

53. others: **Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays**. *Nature biotechnology* 2000, **18**:630–634.

54. Sterky F, Bhalerao RR, Unneberg P, et al. **A Populus EST resource for plant functional genomics**. *PNAS* 2004, **101**:13951-13956.

55. Allona I, Quinn M, Shoop E, et al. **Analysis of xylem formation in pine by cDNA sequencing**. *Plant biology* 1998, **95**:9693-9698.

56. Vinod MS, Sankararamasubramanian HM, Priyanka R, Ganesan G, Parida A: **Gene expression analysis of volatile-rich male flowers of dioecious *Pandanus fascicularis* using expressed sequence tags.** *Journal of plant physiology* 2010, **167**:914-9.

57. Moccia MD, Oger-Desfeux C, Marais GA, Widmer A: **A White Campion (*Silene latifolia*) floral expressed sequence tag (EST) library: annotation, EST-SSR characterization, transferability, and utility for comparative mapping.** *BMC genomics* 2009, **10**:243.

58. Shimamura K, Ishimizu T, Nishimura K, et al. **Analysis of expressed sequence tags from Petunia flowers**. *Plant Science* 2007, **173**:495-500.

59. Shaogui G, Yi Z, Je-Gun J, et al. **Transcriptome sequencing and comparative analysis of cucumber flowers with different sex types**. *BMC Genomics* 2010, **11**.

60. Laitinen R A E, Immanen J, Auvinen P, et al. **Analysis of the floral transcriptome uncovers new regulators of organ determination and gene families related to flower organ differentiation in *Gerbera hybrida* (Asteraceae)**. *Genome research* 2005, **15**:475.

61. Tsai W-C, Hsiao Y-Y, Lee S-H, et al. **Expression analysis of the ESTs derived from the flower buds of *Phalaenopsis equestris***. *Plant Science* 2006, **170**:426-432.

62. Beulé T, Camps C, Debiesse S, et al. **Transcriptome analysis reveals differentially expressed genes associated with the mantled homeotic flowering abnormality in oil palm (*Elaeis guineensis*)**. *Tree Genetics & Genomes* 2010, **7**:169-182.

63. Wu Q, Song J, Sun Y, et al. **Transcript profiles of *Panax quinquefolius* from flower, leaf and root bring new insights into genes related to ginsenosides biosynthesis and transcriptional regulation.** *Physiologia plantarum* 2010, **138**:134-49.

64. Conesa A, Götz S, García-Gómez JM, et al. **Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research.** *Bioinformatics (Oxford, England)* 2005, **21**:3674-6.

65. Altschul SF, Madden TL, Schäffer a a, et al. **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucleic acids research* 1997, **25**:3389-402.

66. Hunter S, Apweiler R, Attwood TK, et al. **InterPro: the integrative protein signature database.** *Nucleic acids research* 2009, **37**:D211-5.

67. Finn RD, Mistry J, Tate J, et al. **The Pfam protein families database.** *Nucleic acids research* 2010, **38**:D211-22.

68. Charbonnel-Campaa L, Lauga B, Combes D: **Isolation of a type 2 metallothionein-like gene preferentially expressed in the tapetum in *Zea mays*.** *Gene* 2000, **254**:199-208.

69. Hunter DA, Steele BC, Reid MS: **Identification of genes associated with perianth senescence in Daffodil (*Narcissus pseudonarcissus* L. Dutch Master)**. *Plant Science* 2002, **163**:13–21.

70. Nordberg M: **Metallothioneins: historical review and state of knowledge.** *Talanta* 1998, **46**:243-54.

71. Miller JD, Arteca RN, Pell EJ: **Senescence-associated gene expression during ozone-induced leaf senescence in Arabidopsis**. *Plant Physiology* 1999, **120**:1015.

72. Etscheid M, Klumper S, Riesner D: **Accumulation of a Metallothionein-like mRNA in Norway Spruce Under Environmental Stress**. *Journal of Phytopathology* 1999, **147**:207–213.

73. Davies C, Robinson SP: **Differential screening indicates a dramatic change in mRNA profiles during grape berry ripening. Cloning and characterization of cDNAs encoding putative cell wall and stress response proteins.** *Plant physiology* 2000, **122**:803-12.

74. Reid S: **Up-regulation of two cDNA clones encoding metallothionein-like proteins in apple fruit during cool storage**. *Physiologia Plantarum* 1997, **100**:183-189.

75. JONES CS, DAVIES HV, MCNICOL RJ, TAYLOR MA: **Cloning of three genes up-regulated in ripening raspberry fruit (Rubus idaeus cv. Glen Clova)**. *Journal of plant physiology* , **153**:643-648.

76. Buchanan-Wollaston V: **Isolation of cDNA clones for genes that are expressed during leaf senescence in Brassica napus. Identification of a gene encoding a senescence-specific metallothionein-like protein**. *Plant Physiology* 1994, **105**:839-846.

77. Buchanan-Wollaston V, Ainsworth C: **Leaf senescence in Brassica napus: cloning of senescence related genes by subtractive hybridisation**. *Plant Molecular Biology* , **33**:821-834.

78. Breeze E, Wagstaff C, Harrison E, et al. **Gene expression patterns to define stages of post-harvest senescence in Alstroemeria petals**. *Plant Biotechnology Journal* 2004, **2**:155-168.

79. Channelière S, Rivière S, Scalliet G, et al. **Analysis of gene expression in rose petals using expressed sequence tags.** *FEBS letters* 2002, **515**:35-8.

80. Swanson R, Clark T, Preuss D: **Expression profiling of Arabidopsis stigma tissue identifies stigma-specific genes**. *Sexual Plant Reproduction* , **18**:163-171.

81. Tung C, Dwyer K, Nasrallah M: **Genome-wide identification of genes expressed in Arabidopsis pistils specifically along the path of pollen tube growth**. *Plant* 2005, **138**:977-989.

82. Goldman M, Pezzotti M, Seurinck J, Mariani C: **Developmental expression of tobacco pistil-specific genes encoding novel extensin-like proteins**. *The Plant Cell Online* 1992, **4**:1041-51.

83. Yamane H, Lee S-J, Kim B-D, Tao R, Rose JKC: **A coupled yeast signal sequence trap and transient plant expression strategy to identify genes encoding secreted proteins from peach pistils.** *Journal of experimental botany* 2005, **56**:2229-38.

84. Quiapim AC, Brito MS, Bernardes L a S, et al. **Analysis of the Nicotiana tabacum stigma/style transcriptome reveals gene expression differences between wet and dry stigma species.** *Plant physiology* 2009, **149**:1211-30.

85. Allen MD, Yamasaki K, Ohme-Takagi M, Tateno M, Suzuki M: **A novel mode of DNA recognition by a beta-sheet revealed by the solution structure of the GCC-box binding domain in complex with DNA.** *The EMBO journal* 1998, **17**:5484-96.

86. Schopfer P: **Hydroxyl radical-induced cell-wall loosening in vitro and in vivo: implications for the control of elongation growth**. *The Plant Journal* 2001, **28**:679-688.

87. Brownleader MD, Hopkins J, Mobasheri A, et al. **Role of extensin peroxidase in tomato (Lycopersicon esculentum Mill.) seedling growth**. *Planta* , **210**:668-676.

88. Mudge SR, Rae AL, Diatloff E, Smith FW: **Expression analysis suggests novel roles for members of the Pht1 family of phosphate transporters in Arabidopsis.** *The Plant journal : for cell and molecular biology* 2002, **31**:341-53.

89. Chapin LJ, Jones ML: **Ethylene regulates phosphorus remobilization and expression of a phosphate transporter (PhPT1) during petunia corolla senescence.** *Journal of experimental botany* 2009, **60**:2179-90.

90. Schneider S, Schneidereit A, Udvardi P, et al. **Arabidopsis INOSITOL TRANSPORTER2 mediates H+ symport of different inositol epimers and derivatives across the plasma membrane.** *Plant physiology* 2007, **145**:1395-407.

91. Roberts J: **Cell Separation Processes in Plants—Models, Mechanisms and Manipulation**. *Annals of Botany* 2000, **86**:223-235.

92. Sexton R, Roberts JA: **Cell Biology of Abscission**. *Annual Review of Plant Physiology* 1982, **33**:133-162.

93. Abeles F: **Role of RNA and protein synthesis in abscission**. *Plant Physiology* 1968, **43**:1577.

94. Roberts J: **Cell Separation Processes in Plants—Models, Mechanisms and Manipulation**. *Annals of Botany* 2000, **86**:223-235.

95. Bonghi C, Rascio N, Ramina A, Casadoro G: **Cellulase and polygalacturonase involvement in the abscission of leaf and fruit explants of peach**. *Plant Molecular Biology* 1992, **20**:839-848.

96. Lashbrook CC, Gonzalez-Bosch C, Bennett a B: **Two divergent endo-beta-1,4-glucanase genes exhibit overlapping expression in ripening fruit and abscising flowers.** *The Plant cell* 1994, **6**:1485-93.

97. Kalaitzis P, Koehler SM, Tucker ML: **Cloning of a tomato polygalacturonase expressed in abscission**. *Plant Molecular Biology* 1995, **28**:647-656.

98. TAYLOR JE, WEBB STJ, COUPE SA, TUCKER GA, ROBERTS JA: **Changes in Polygalacturonase Activity and Solubility of Polyuronides during Ethylene-stimulated Leaf Abscission in Sambucus nigra** . *Journal of Experimental Botany* 1993, **44** :93-98.

99. Woollard A a D, Moore I: **The functions of Rab GTPases in plant membrane traffic.** *Current opinion in plant biology* 2008, **11**:610-9.

100. Vernoud V, Horton AC, Yang Z, Nielsen E: **Analysis of the Small GTPase Gene Superfamily of Arabidopsis** . *Plant Physiology* 2003, **131** :1191-1208.

101. Cheung AY, Chen CY, Tao L, et al. **Regulation of pollen tube growth by Rac-like GTPases** . *Journal of Experimental Botany* 2003, **54** :73-81.

102. Hwang J-U, Yang Z: **Small GTPases and Spatiotemporal Regulation of Pollen Tube Growth**. In *The Pollen Tube*. edited by Malhó R Springer Berlin / Heidelberg; 2006, **3**:95-116.

103. Kotchoni SO, Jimenez-Lopez JC, Gao D, et al. **Modeling-Dependent Protein Characterization of the Rice Aldehyde Dehydrogenase (ALDH) Superfamily Reveals Distinct Functional and Structural Features**. *PLoS ONE* 2010, **5**:e11516.

104. Wei Y, Lin M, Oliver DJ, Schnable PS: **The roles of aldehyde dehydrogenases (ALDHs) in the PDH bypass of Arabidopsis.** *BMC biochemistry* 2009, **10**:7.

105. Zielinski RE: **Calmodulin and calmodulin-binding proteins in plants**. *Annual Review of Plant Biology* 1998, **49**:697–725.

106. Arazi T, Baum G, Snedden WA, Shelp BJ, Fromm H: **Molecular and Biochemical Analysis of Calmodulin Interactions with the Calmodulin-Binding Domain of Plant Glutamate Decarboxylase** . *Plant Physiology* 1995, **108** :551-561.

107. Bachs O, Agell N, Carafoli E: **Calmodulin and calmodulin-binding proteins in the nucleus**. *Cell Calcium* 1994, **16**:289-296.

108. Raz V, Fluhr R: **Ethylene Signal Is Transduced via Protein Phosphorylation Events in Plants** . *The Plant Cell Online* 1993, **5** :523-530.

109. Abeles F: **Role of RNA and protein synthesis in abscission**. *Plant Physiology* 1968, **43**:1577.

110. Takatsuji H: **Zinc-finger transcription factors in plants.** *Cellular and molecular life sciences : CMLS* 1998, **54**:582-96.

111. Weigel D: **From floral induction to floral shape**. *Current Opinion in Plant Biology* 1998, **1**:55–59.

112. Hu W, Wang Y, Bowers C, Ma H: **Isolation, sequence analysis, and expression studies of florally expressed cDNAs in Arabidopsis**. *Plant Molecular Biology* 2003, **53**:545-563.

113. Zhao D, Yu Q, Chen C, Ma H: **Genetic control of reproductive meristems**. *Sheffield Academic Press* 2001:89– 142.

114. Santos LA: **Efeito da Superexpressão dos Fatores de Transcrição ZmDof1 e OsDof25 sobre a Eficiência de Uso de Nitrogênio em *Arabidopsis thaliana* L.** *Carbon* 2009.

115. Zhang J: **Overexpression analysis of plant transcription factors**. *Current Opinion in Plant Biology* 2003, **6**:430-440.

116. Liu L, White MJ, Macrae TH: **Functional domains , evolution and regulation**. 1999, **257**.

117. Cero JD: **Análise "in silico" e de expressão da família gênica ETHYLENE RESPONSIVE FACTORS (ERF) no gênro Malus**. 2010.

118. Perez-Rodriguez P, Riano-Pachon D, Correa LGG, et al. **PlnTFDB: updated content and new features of the plant transcription factor database**. *Nucleic Acids Research* 2009.

119. Guo AY, Chen X, Gao G, et al. **PlantTFDB: a comprehensive plant transcription factor database**. *Nucleic Acids Research* 2008, **36**:D966.

120. Dietz K-J, Vogel MO, Viehhauser A: **AP2/EREBP transcription factors are part of gene regulatory networks and integrate metabolic, hormonal and environmental signals in stress acclimation and retrograde signalling.** *Protoplasma* 2010:3-14.

121. Feng J-X, Liu D, Pan Y, et al. **An annotation update via cDNA sequence analysis and comprehensive profiling of developmental, hormonal or environmental responsiveness of the Arabidopsis AP2/EREBP transcription factor gene family.** *Plant molecular biology* 2005, **59**:853-68.

122. Huibers RP, Jong M de, Dekter RW, Van den Ackerveken G: **Disease-specific expression of host genes during downy mildew infection of Arabidopsis.** *Molecular plant-microbe interactions : MPMI* 2009, **22**:1104-15.

123. Balazadeh S, Riaño-Pachón DM, Mueller-Roeber B: **Transcription factors regulating leaf senescence in *Arabidopsis thaliana*.** *Plant biology (Stuttgart, Germany)* 2008, **10 Suppl 1**:63-75.

124. Yamasaki K, Kigawa T, Inoue M, Tateno M, Yamasaki T: **Solution Structure of the B3 DNA Binding Domain of the Arabidopsis Cold-Responsive Transcription Factor RAV1**. *Structure* 2004, **16**:3448-3459.

125. Kagaya Y, Ohmiya K, Hattori T: **RAV1, a novel DNA-binding protein, binds to bipartite recognition sequence through two distinct DNA-binding domains uniquely found in higher plants**. *Nucleic acids research* 1999, **27**:470.

126. Swaminathan K, Peterson K, Jack T: **The plant B3 superfamily.** *Trends in plant science* 2008, **13**:647-55.

127. GUO Y, CAI Z, GAN S: **Transcriptome of Arabidopsis leaf senescence**. *Plant, Cell & Environment* 2004, **27**:521-549.

128. Wellmer F, Riechmann JL, Alves-Ferreira M, Meyerowitz EM: **Genome-wide analysis of spatial gene expression in arabidopsis flowers**. *Plant Cell* 2004, **16**:1314-1326.

129. Hennig L, Gruissem W, Grossniklaus U, Köhler C: **Transcriptional programs of early reproductive stages in arabidopsis**. *Plant Physiology* 2004, **135**:1765-1775.

130. Seki M, Narusaka M, Ishida J, et al. **Monitoring the expression profiles of 7000 Arabidopsis genes under drought, cold and high-salinity stresses using a full-length cDNA microarray**. *The Plant Journal* 2002, **31**:279-292.

131. Hoth S, Morgante M, Sanchez J-P, et al. **Genome-wide gene expression profiling in *Arabidopsis thaliana* reveals new targets of abscisic acid and largely impaired gene regulation in the abi1-1 mutant** . *Journal of Cell Science* 2002, **115** :4891-4900.

132. Schenk PM, Kazan K, Manners JM, et al. **Systemic gene expression in Arabidopsis during an incompatible interaction with *Alternaria brassicicola***. *Plant Physiology* 2003, **132**:999-1010.

133. Oono Y, Seki M, Nanjo T, et al. **Monitoring expression profiles of Arabidopsis gene expression during rehydration process after dehydration using ca. 7000 full-length cDNA microarray**. *The Plant Journal* 2003, **34**:868-887.

134. Rabbani MA, Maruyama K, Abe H, et al. **Monitoring Expression Profiles of Rice Genes under Cold, Drought, and High-Salinity Stresses and Abscisic Acid Application Using cDNA Microarray and RNA Gel-Blot Analyses** . *Plant Physiology* 2003, **133** :1755-1767.

135. Deslandes L, Olivier J, Theulières F, et al. **Resistance to *Ralstonia solanacearum* in *Arabidopsis thaliana* is conferred by the recessive RRS1-R gene, a member of a novel family of resistance genes** . *Proceedings of the National Academy of Sciences* 2002, **99** :2404-2409.

136. Dellagi A, Heilbronn J, Avrova AO, et al. **A Potato Gene Encoding a WRKY-like Transcription Factor Is Induced in Interactions with *Erwinia carotovora subsp. atroseptica* and *Phytophthora infestans* and Is Coregulated with Class I Endochitinase Expression**. *Molecular Plant-Microbe Interactions* 2000, **13**:1092-1101.

137. Dong J, Chen C, Chen Z: **Expression profiles of the *Arabidopsis* WRKY gene superfamily during plant defense response**. *Plant Molecular Biology* , **51**:21-37.

138. Du L, Chen Z: **Identification of genes encoding receptor-like protein kinases as possible targets of pathogen- and salicylic acid-induced WRKY DNA-binding proteins in Arabidopsis**. *The Plant Journal* 2000, **24**:837-847.

139. Chen C, Chen Z: **Potentiation of Developmentally Regulated Plant Defense Response by AtWRKY18, a Pathogen-Induced Arabidopsis Transcription Factor** . *Plant Physiology* 2002, **129** :706-716.

140. Chen W, Provart NJ, Glazebrook J, et al. **Expression Profile Matrix of Arabidopsis Transcription Factor Genes Suggests Their Putative Functions in Response to Environmental Stresses** . *The Plant Cell Online* 2002, **14** :559-574.

141. Wang Z, Yang P, Fan B, Chen Z: **An oligo selection procedure for identification of sequence-specific DNA-binding activities associated with the plant defence response**. *The Plant Journal* 1998, **16**:515-522.

142. Yang P, Chen C, Wang Z, Fan B, Chen Z: **A pathogen- and salicylic acid-induced WRKY DNA-binding activity recognizes the elicitor response element of the tobacco class I chitinase gene promoter**. *The Plant Journal* 1999, **18**:141-149.

143. Beyer K, Binder A, Boller T, Collinge M: **Identification of potato genes induced during colonization by Phytophthora infestans**. *Molecular Plant Pathology* 2001, **2**:125-134.

144. Kalde M, Barth M, Somssich IE, Lippok B: **Members of the Arabidopsis WRKY Group III Transcription Factors Are Part of Different Plant Defense Signaling Pathways**. *Molecular Plant-Microbe Interactions* 2003, **16**:295-305.

145. Cheong YH, Chang H-S, Gupta R, et al. **Transcriptional Profiling Reveals Novel Interactions between Wounding, Pathogen, Abiotic Stress, and Hormonal Responses in Arabidopsis** . *Plant Physiology* 2002, **129** :661-677.

146. Rizhsky L, Liang H, Mittler R: **The Combined Effect of Drought Stress and Heat Shock on Gene Expression in Tobacco** . *Plant Physiology* 2002, **130** :1143-1151.

147. Pnueli L, Hallak-Herr E, Rozenberg M, et al. **Molecular and biochemical mechanisms associated with dormancy and drought tolerance in the desert legume *Retama raetam***. *The Plant Journal* 2002, **31**:319-330.

148. Huang T, Duman JG: **Cloning and characterization of a thermal hysteresis (antifreeze) protein with DNA-binding activity from winter bittersweet nightshade, *Solanum dulcamarat*;**. *Plant Molecular Biology* , **48**:339-350.

CAPÍTULO II

Comparative Phylogeny and Expression Analysis of Genes Encoding Dof-type Transcription Factors from *Eucalyptus grandis*

(Manuscrito submetido ao periódico *BMC Plant Biology*)

Comparative Phylogeny and Expression Analysis of Genes Encoding Dof-type Transcription Factors from *Eucalyptus grandis*

Michèle C Breton[1], Gabriel S D´Almeida[1], Sandro S Camargo[2], Luís F Revers[3], Jeverson Frazzon[1,4] , Giancarlo Pasquali[1§]

[1]Graduate Program in Cell and Molecular Biology, Centro de Biotecnologia, Universidade Federal do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, Brazil

[2]Universidade Federal do Pampa, Campus Bagé, Travessa 45, 1650, sala 2107, Bairro Malafaia, CEP 96413-170, Bagé, RS, Brazil

[3]Laboratório de Biologia Molecular Vegetal, Embrapa Uva e Vinho, Bento Gonçalves, RS, Brazil

[4]Instituto de Ciência e Tecnologia de Alimentos, Universidade Federal do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, Brazil

[§]Corresponding author: Centro de Biotecnologia, Universidade Federal do Rio Grande do Sul, Av. Bento Gonçalves, 9500, Bairro Agronomia, Caixa Postal: 15053, CEP: 91501-970, Porto Alegre, Rio Grande do Sul, Brazil

Email addresses:

MCB: mbreton@uol.com.br

GSD: gsdalmeida@gmail.com

SSC: sandro.camargo@unipampa.edu.br

LFR: luis@cnpuv.embrapa.br

JF: jeverson.frazzon@ufrgs.br

GP: pasquali@cbiot.ufrgs.br

Abstract

Background

Dof proteins are a family of transcription factors specific to the plant kingdom that contain a particular class of zinc finger DNA binding domain. Members of this family are involved in the regulation of genes related to a plethora of metabolic processes including stress or hormone response and cell or tissue specificity. Dof proteins and encoding genes were characterized in several plant species. Nevertheless poplar is the only woody species whose *Dof* genes were better characterized. The recent availability of the *Eucalyptus grandis* genome and transcriptome, along with transcription factor databases of several plant species allowed us to identify and run a valuable comparative analysis of the Dof protein family in this tree. The phylogenetic relationships among Dof proteins from *E. grandis* and *Arabidopsis thaliana* is a fundamental step to unravel functionality of new *Dof* genes not yet characterized.

Results

Twenty-three distinct DNA sequences were predicted to belong to the *Dof* gene family after the analysis of the complete available genome of *E. grandis*. The deduced protein sequences of 22 members do contain a conserved Dof domain. One sequence seemed to have lost the conserved Dof domain, suggesting it to be a pseudogene or with an activity not directly linked to the Dof family. Gene structures, including exon/intron positions, and amino acid sequences were predicted for each gene based on the available *E. grandis* transcriptome. In order to determine the relationship and function of the genes putatively encoding Dof proteins, we carried out a phylogenetic analysis with 43 Dof protein sequences from *A. thaliana*. Our analysis allowed us to classify the *E. grandis Dof* sequences into five groups of orthologous genes. Gene expression analysis via real time, quantitative PCR was also conducted with ten of the *E. grandis Dof* genes. Generally, *Dof* steady-state mRNA levels were higher in *E. grandis* vascular tissues, with more reduced levels in flowers. *Dof* genes showed a general increase in

steady-state mRNA levels after hormone signaling, and general reduced levels following abiotic stress.

Conclusions

This is the first study that aimed the identification of *Dof* genes in *E. grandis* which are possibly involved in numerous plant metabolic processes. The phylogenetic relationship to *A. thaliana* counterparts and the patterns of mRNA accumulation in *E. grandis* allowed us to speculate on possible roles for some of the Dof-encoding genes.

Background

In Brazil, the area of *Eucalyptus* planted forests is rapidly expanding, with an average area increase of 7.1% per year between 2004 and 2009, totaling 3.5 million hectares last year. This expansion is due to a series of factors including the rapid growth of *Eucalyptus* trees allowing short rotation cycles, the high forest productivity in Brazil and some other countries, and the expansion and direction of new investments by the business segment that use wood as raw material in industrial processes. The use of new technologies at the industrial scale enabled a new opportunity to the forestry sector, providing advantages in the competitive global market. The improvement of agronomic traits by biotechnology is already a widely employed and valuable tool in agriculture and becomes available for forest species, such as *Eucalyptus* [195].

Advances in the knowledgement of plant genomics and biological systems has been a valuable tool to identify regulatory genes and metabolic pathways that control important characteristics. Acting as key regulators of cellular processes, transcription factors become excellent candidates for modifying complex traits in plants [196] and potential candidates for achieving success in biotechnology associated with agroforestry.

The family of Dof (DNA one finger) transcription factors, exclusively present in plants, has been associated with several functions in gene regulation and specific processes in plant metabolism. Many Dof proteins have been described both in dicotyledonous and monocotyledonous plants. Currently only a small fraction of these proteins have their involvement in plant metabolism described [169].

Biochemical and molecular genetics analysis has shown that the Dof DNA binding domain is a zinc finger of the Cys_2-Cys_2 type, including 50 amino acids that specifically bind DNA sequences with the 5'-T/AAAAG-3' core. The conserved Dof domain, which is normally present in a single copy at the N-terminal protein region, is able to confer specificity of binding to DNA. However regions outside the Dof domain

can be quite diverse and are known to be linked to several different roles played by these proteins [166, 197]. All Dof domains, with their four conserved cysteine residues, have peculiar amino acid sequences and an arrangement that is different from other zinc finger motifs [169, 170]. The replacement of the conserved cysteine residues by alanine or serine completely inhibited the binding of Dof domain proteins to DNA *in vitro* [198] and *in vivo* [180], proving their essentiality.

The Dof proteins act as transcription activators or repressors in several metabolisms like stress-response [168], plant defense mechanism [169], light regulation of genes involved in carbohydrate metabolism [170-172], phytochrome signaling [173], response to plant hormones including auxin [174, 175], and gibberellins[176,177], seed germination [178, 179], seed storage protein synthesis in developing endosperm [180], and stomata guard cell specific gene regulation [181].

Currently, many theoretical and analytical tools of systems phylogenetics have been used in the field of comparative genomics, since the hierarchical organization of genes reflects old processes of divergence, duplication and/or deletion of genes [200], aiding in the inference of the biological role of each protein arising out of these genes along the evolution.

In order to better understand the importance of *Dof* genes and encoded proteins in plants of agroforestry interest such as *Eucalyptus*, especially with regard to responses to biotic and abiotic stresses, we started an effort to characterize this family of transcription factors. In this study, we describe the identification of 22 gene sequences containing the complete Dof domain, based on data mining of whole *E. grandis* genome and transcriptome. We conducted a phylogenetic analysis in order to predict roles possibly played by each of the putatively encoded Dof proteins from *E. grandis* and the definition of orthologous groups along with Dof protein sequences from *A. thaliana*. Finally, the expression of ten putative *E. grandis Dof* genes was assessed by reverse transcription followed by quantitative, real-time polymerase chain reaction (RT-qPCR).

Results and Discussion

Identification and determination of the genetic architecture of Dof transcription factors in *Eucalyptus grandis*

We analyzed the available *E. grandis* genome (Phytozome, http://www.phytozome.net) and transcriptome (GENOLYPTUS Database; http://genoma.embrapa.br/genoma/genolyptus) of *E. grandis* in order to generate a set of non-redundant sequences containing Dof domains. Dof domain nucleotide sequences from 38 plant species, including two gymnosperms, nine monocots and 27 dicots were used as templates for searching the databases with the Basic Local Alignment Search Tool (BLAST) [65]. *Dof* nucleotide sequences were extracted from the Plant Transcription Factor Databases of the University of Potsdam, (Germany) [118] and of the Peking University (China) [119]. Database searches resulted in 23 nucleotide scaffolds of the *E. grandis* genome showing significant alignments with Dof domain templates. However, one of the sequences exhibited an incomplete Dof domain, with many amino acids absent including two key cysteines (result not shown). We suggest that this sequence is possibly a pseudogene or that its activity is not directly linked to the Dof family of transcription factors.

The *E. grandis Dof*-containing nucleotide scaffolds were subjected to a self-training algorithm to predict the presence of open reading frames (ORFs) in eukaryotic genomes, the GeneMark.hmm ES-3.0 [201]. Resulting amino acid sequences predicted as belonging to the Dof family of transcription factors were confirmed by protein BLAST of the databases available at the National Center for Biotechnology Information (NCBI) [65] and The Arabidopsis Information Resource (TAIR) [221], followed by the identification of the Dof protein domain by Pfam [202] (Table 1). These data were integrated with the GENOLYPTUS database of transcript sequences, in order to confirm the expressed identity of the genes. As a result, 22 out of the 23 sequences exhibited perfect ORFs with conserved Dof domains. The genomic, cDNA and amino acid

sequences of *E. grandis* putative Dof transcription factor are presented in the Additional file 1.

The analysis of the *E. grandis* genomic scaffolds selected using the Artemis Release 12.0 tool [203] allowed us to build a probable architecture of the predicted *Dof* sequences, as shown in Figure 1. The direction of gene transcription in each scaffold is represented by the arrows in Figure 1. Typically, most Dof-containing sequences exhibited one (*EucaDof7*, *8*, *13*, *14*, *16* and *18*) or two exons (*EucaDof1*, *2*, *3*, *5*, *10*, *11*, *20*, *21* and *22*), totaling 15 sequences. Remaining seven sEaffolds contained three (*EucaDof12*, *15* and *17*), four (*EucaDof4*), five (*EucaDof6* and *9*) or six exons (*EucaDof19*). Exon sequences varied from 264 bp (*EucaDof21*) up to 2,331 bp (*EucaDof19*), while introns ranged from 25 bp up to 4,700 bp, demonstrating the high size variability of the putative *E. grandis Dof* genes. We observed that 63.6% of the sequences have the sense direction of transcription, while 36.4% are counter-intuitive.

Phylogenetic analysis of Dof protein sequences from *A. thaliana* and *E. grandis*

The predicted amino acid sequences of the Dof domains derived from the 22 putative *E. grandis Dof* genes were aligned in order to evaluate the evolutionary relationship among them (Figure 2). As expected, all amino acid residues that are known to be highly conserved among the described DNA-binding Dof domains were perfectly mapped in all 22 *E. grandis Dof* genes.

Based on the phylogenetic analysis conducted by Lijavetzky et al. with the Dof-deduced protein families from *A. thaliana* and rice (*Oryza sativa indica* and *japonica*) [197], we performed a similar comparison with the Dof proteins from *E. grandis* and *A. thaliana* [204]. Initially, to determine the conservation of amino acid residues and the presence of other domains in *E. grandis* Dof-encoded proteins, we employed the Multiple EM Motif Elicitation Tool (MEME) [205] fit to the same parameters used by Lijavetzky et al. [197]. The MEME analysis allowed us to predict 36 protein domains besides the Dof domain as shown in Figure 3 and in the Table 2.

The result of the multiple alignment of the total set of 22 deduced proteins from *E. grandis* (Figure 3) with 43 proteins from *A. thaliana* revealed that the Dof spanning region was the only one shared among all the sequences. In total, 13 of the 37 motifs identified were shared between species Table 2. Among the motifs, we observed that motif 31 was the only one almost exclusive to *E. grandis* sequences, found in proteins EucaDof3, 6, 11 and 19. Only the *A. thaliana* At3g47500.1 deduced protein, within the same group, shared the same motif.

To evaluate the evolutionary relationships among Dof proteins, we performed a combined phylogenetic analysis with the 65 sequences of *A. thaliana* and *E. grandis* using the Molecular Evolutionary Genetics Analysis (MEGA) v.4.1 tool [206] and the resulting data was converted into a phylogenetic graphic by FigTree v1.3.1. (http://tree.bio.ed.ac.uk/). The unrooted tree inferred by the Neighbor-Joining method allowed us to identify four major clusters of orthologous groups (MCOGs) called A, B, C and D (Figure 4). The MCOGs were further divided into subgroups (B1 and B2; C1, C2.1, C2.2 and C3) supported by the boot-strapping values and the occurrence of common protein motifs outside the Dof domain, where several motifs were indeed found to be commonly linked (e.g. 24-12-23-25, 37-30 and 16-35-37-10-5-20). The resulting analysis of individual *E. grandis* Dof protein sequences and the tree topology, as well as their organization, was similar to that described for *A. thaliana*.

As shown in Figure 4, putative *E. grandis* Dof proteins have representatives in all groups and subgroups described for *A. thaliana* [16], clearly representing possible orthologous genes. Genes predicted to be orthologs may retain similar functions in cellular metabolism while paralogs usually have different functions [200]. Five *E. grandis* Dof-deduced proteins exhibited very high similarities to *A. thaliana* counterparts, as can be seen for EucaDof7, 11, 16, 21 and 22 in Figure 4, making them the best orthologous candidates. Among putative paralogous sequences, EucaDof1 and 14, 5 and 17, 8 and 9, 10 and 15, and 12 and 20 were the closest sequences within *E. grandis*, possibly representing events of gene duplication.

When comparing multigene families between species, it is common to observe several genes of a species that are collectively orthologous to a single gene of another, indicating recent duplications unique to the first. Since the orthologous or

paralogous relationship among genes is difficult to infer taking into account only the topology of a tree, prior knowledge about the function of some of the genes would allow the confirmation of such a condition. This is the case for the DAG1 (At3g61850.1, At3g61850.2, At3g61850.4) and DAG2 (At2g46590.1, At2g46590.2) genes previously described in *A. thaliana*, and that are closely related to EucaDof4 in subgroup C2.1, as observed in Figure 4 and Table 1. DAG1 and DAG2 were shown to exhibit a high degree of similarity between sequences and to have identical patterns of expression, indicating a potential case of functional redundancy. However, these genes were proven to have opposite functions in controlling seed germination after a systematic analysis of *A. thaliana* mutant strains [178, 179, 197]. EucaDof4 is the most similar sequence to DAG2 found in *E. grandis* and its function in seed germination has still to be determined. It has an insertion of 95 amino acid residues in the C-terminal protein sequence as compared to DAG1 and DAG2 of *A. thaliana*. Thus, DAG1 and DAG2 genes are paralogous produced by a recent duplication event in the genome of *A. thaliana*, whose standard was not maintained in *E. grandis*.

Evaluation of *E. grandis Dof* gene spatial expression by RT-qPCR

Ten out of the 22 gene sequences belonging to the putative Dof transcripton factor family of *E. grandis* were selected to evaluate their spatial expression profile by RT-qPCR. The first evaluation was made with respect to the expression pattern presented in plant organs, whereas the second approach was intended to evaluate *Dof* gene expression in seedlings subjected to treatments that mimic hormonal cell signals or abiotic stresses. Therefore for the first set of assays, total RNA was extracted from flowers, leaves and vascular tissues (xylem) of *E. grandis* adult trees. Total RNA samples were subsequently treated with DNase I and submitted to first strand cDNA synthesis prior to RT-qPCRs. Conventional PCRs followed by agarose gel electrophoresis and amplicon sequencing were performed with all primer pairs, confirming the identity of all amplified fragments obtained (results not shown).

Primers for RT-qPCR were designed for *EucaDof2*, *3*, *4*, *6*, *10*, *12*, *13*, *17*, *21* and 22 (Additional file 2). Three housekeeping genes known to exhibit constitutive expression in other plant systems [38] were chosen to normalize the values of steady-state mRNA levels measured by RT-qPCR: ribosomal protein L23A (RibL23A), aspartyl-tRNA synthetase (*Eucons08*) and transcription elongation factor S-II (*Eucons21*). The set of primers designed for these *E. grandis* sequences are presented in the Additional file 2. All RT-qPCR measurements were performed with biological duplicates and techinical quadruplicates.

The expression profiles of the selected *E. grandis Dof* genes obtained by RT-qPCR analysis were compared to available expression profiles generated by microarrays for *A. thaliana* using the Genevestigator visualization tool [208]. Genevestigator is a system built to graphically represent gene expression and regulation in different species, taking into consideration various factors such as tissue type, disease, genetic modifications or external stimuli. Its methodology is based on a query to the database manually prepared that combines results from tens of thousands of microarrays, which allows various types of advanced analysis, as the search for expression profiles under certain physiological conditions or certain tissues [208].

Although we have considered the expression profiles of *A. thaliana Dof* genes generated by microarray analysis as a parameter of comparison for our RT-qPCR results, one should not expect good correlation between the signals obtained from microarray and the level of individual transcripts by RT-qPCR. Microarrays are based on the hybridization of probes and, therefore, it is a qualitative technique, while RT-qPCR is a quantitative method. Hence it is important to consider that there is no strictly linear relationship between signal strength found in microarrays and the amount of transcripts for different genes assessed by RT-qPCR during the analysis [209, 210].

In order to define the best normalizing organ/tissue to which other organ/tissue expression values could be compared, we analyzed the Genevestigator expression profile of *A. thaliana Dof* genes, as shown in Figure 5A. It was observed that *A. thaliana Dof* gene expression profiles displayed by flower and leaf were very similar, with a *p-value* between 0.04 and 0.06 and, in some cases, such as that for *At2g28510,*

At4g21050 and *At1g21340*, there was no signal detected (*p>0.06*). In xylem, a detectable signal was only observed for genes *At3g47500* and *At4g24060*, with a *p-value* lower than 0.04 [208]. Therefore, we decided to choose leaf as the reference organ for normalization and comparison of gene expressions measured by RT-qPCR.

According to our phylogenetic analysis (Figure 4), the EucaDof3 protein sequence is orthologous to *A. thaliana* At3g47500.1 and At5g62430.1 proteins, and paralogous to the EucaDof6 sequence. As shown in Figure 5B and 5C, both *EucaDof3* and *EucaDof6* exhibited higher levels of steady-state mRNA in xylem, followed by leaves and flowers. This result is in agreement with the microarray-measured pattern of expression of the *A. thaliana* orthologs At3g47500.1 and At5g62430.1, as shown in Figure 5A. Proteins encoded by *At5g62430.1* and *At3g47500.1*, respectively CDF1 and CDF3, are known to be involved in the regulation of the transition time between plant vegetative and reproductive phases. They do so by thereby stabilizing the Flavin-binding, Kelch repeat, F-box 1 (FKF1) protein that, in turn, regulates the transcription of CONSTANS, a protein involved in the photoperiodic flowering mechanism in *A. thaliana* [211]. Additionally, according to results obtained by overexpressing CDF1 (*At5g62430.*1) or CDF3 (*At3g47500.*1) in plants, *cdf1-mutant* flowered later than wild-type in long-day conditions. In contrast, the *cdf3-mutant* did not change its flowering time, independently of the length of the day. This raised the possibility that FKF1 might be responsible for regulating the CDF1 protein turnover *in vivo* [33]. Besides the close homology at the protein sequence level, the expression pattern of the *A. thaliana* CDF1-enconding gene (*At5g62430.*1), observed mainly in the vascular tissues of cotyledons, leaves and hypocotyls, as well in stomata [211], reinforces the possibility of a similar function for *EucaDof3* and *6* in *E. grandis*.

The EucaDof4-encoded protein was grouped within MCOG C2.1 group of *A. thaliana* proteins, without any obvious orthologous or paralogous protein (Figure 4). Its spatial expression in *E. grandis* was shown to be identical to *EucaDof3* and *6*, with higher expression in xylem followed by leaves and flowers, although not as clearly defined as the last two genes.

The paralogous protein sequences encoded by *EucaDof10* and *15* are orthologous to At2g37590.1 and At5g02460.1 (Figure 4). In the RT-qPCR analysis, the

expression of *EucaDof10* was similar in xylem and leaves, being much lower in flowers (Figure 5B and 5C). The microarray expression profile observed for At5g02460.1 differed from the pattern of *EucaDof10* expression, being similar in flowers and leaves and slightly lower in xylem (Figure 5A). The At5g02460.1-encoded protein is involved in the regulation of adaxial and abaxial polarity, by connecting itself directly to the promoter region of the *REVOLUTA* gene (REV) whose expression confers polarity to plant structures. *A. thaliana* mutants for *At5g02460*.1 showed decreased levels of transcription of the *IAA6* and *IAA19* genes, indicating a change in auxin biosynthesis [212]. The observation that both *EucaDof10* and *At5g02460.1* are more significantly expressed in leaves may indicate a possible role of the encoded *E. grandis* transcription factor in organ polarity, like its *A. thaliana* counterpart.

EucaDof12, which exhibited a pattern of expression similar to *EucaDof10* (Figure 5B and 5C), encodes a protein sequence that is paralogous to EucaDof20 and orthologous to At2g28510.1, At5g60200.1 and At3g45610.1 (Figure 4). The microarray expression profile of At2g28510.1 is similar in all organs analyzed (Figure 5A). This *A. thaliana* protein is known to be involved in the formation of leaf vascular networks, being consistently associated with recurring steps in vein ontogeny [213]. Again, the significant expression of *EucaDof12* in leaves may suggest a similar function in *E. grandis*.

EucaDof2-encoded protein is orthologous to *A. thaliana* proteins encoded by At1g28310.1, At1g28310.2, At4g38000.1 and At5g65590.1 (Figure 4). While *At1g28310.1* and *At1g28310.2* have no known functions, *At5g65590.1* encodes a protein that was first observed in stage 2 floral meristem and its expression appeared to continue exclusively in developing sepals at later stages of development [214]. *At4g38000.1* encodes a protein known to participate in the control of floral organ abscission, being part of the transcription complex that directly regulates the expression of cell wall hydrolytic enzymes [215]. While the microarray expression profile of the *At4g38000.1* gene is similar for all three organs studied (Figure 5A), the RT-qPCR expression of *EucaDof2* showed higher levels in leaves followed by xylem and flowers (Figure 5B and 5C). Since flowers employed in this study were at their full

blooming, it is reasonable to believe that *EucaDof2* is not yet activated, assuming that its product has a function similar to the protein encoded by the At4g38000.1 gene.

In our analysis by RT-qPCR, *EucaDof17* showed similar expression values in flowers, xylem, and leaves (Figure 5B and 5C). EucaDof17 is the paralog of EucaDof5. According to Figure 4, the closest *A. thaliana* protein sequence is At4g00940.1, with no known function presented yet.

According to our phylogenetic analysis, the protein sequence encoded by *EucaDof21* is the perfect ortholog to the *A. thaliana* At1g21340.1, whose function is also unknown and therefore no functional speculation could be drawn. Additionally, the comparison of the expression profiles exhibited in microarrays by *At1g21340* (Figure 5A) and in RT-qPCR by *EucaDof21* (Figure 5B and C) demonstrates that transcript levels are discordant. While expression values are similar in all *A. thaliana* organs described for *At1g21340*, we observed decreased *EucaDof21* transcript levels in flowers and xylem, when comparing with values present in leaves.

EucaDof22 protein sequence is the perfect ortholog to At1g64620.1 and close to At4g24060.1 (Figure 4). Again, the roles of the *A. thaliana* proteins have not yet been established. The microarray expression profile of *At4g24060.1* transcripts, as shown in Figure 5A, exhibited a minimum level of expression in leaves followed by flowers, and xylem (highest value observed). The results obtained by RT-qPCR showed that the expression values for *EucaDof22* were the lowest in flowers, with higher and similar levels in leaves and xylem (Figure 5B and 5C). Again, the patterns of transcript accumulation for these orthologous genes are discordant.

The EucaDof13 protein sequence is phylogenetically grouped with *A. thaliana* MCOG C3 proteins whose pattern of domains are very similar to each other [197], but only with the Dof domain and the domain 2 in common with the *E. grandis* protein. The expression profile of the *At4g21050.1* gene by microarray is identical among organs (Figure 5A), while the *EucaDof13* RT-qPCR analysis showed a high level of transcripts in xylem with lower and similar levels in flowers and leaves (Figure 5B and 5C). The *At4g21050.1*-encoded protein is one of the three Dof transcription factors known to be expressed in subsets of cells in the developing epidermis, indicating that it may be involved in cell-specific differentiation and root hair formation

[216]. The relatively distant similarity between EucaDof13 and the *A. thaliana* proteins (Figure 4) and the discordant pattern of expression observed do not allowed us to draw a functional correlation.

Evaluation of *E. grandis Dof* gene expression in response to hormonal or stress by RT-qPCR

Considering that each species may respond differently to biotic and abiotic stimuli, especially when dealing with woody *versus* herbaceous species, a low correlation between expression profiles is expected. Therefore, we limited ourselves to further show the different levels of expression of the *E. grandis Dof* genes in response to hormonal and abiotic stimuli. Ten out of the 22 sequences belonging to the putative Dof transcripton factor gene family of *E. grandis* were further evaluated by RT-qPCR after exposure of *in vitro* grown seedlings to different hormonal and stress-related signals. Total RNA was extracted from three-month-old *in vitro* cultured seedlings subjected to abscisic acid (ABA), 1-naphthaleneacetic acid (NAA) or kinetin (KIN) treatments or abiotic stresses imposed by sodium chloride (NaCl), cold or drought. RNA samples were also extracted from untreated (control) and water-treated seedlings. Total RNA samples were subsequently treated with DNase I and submitted to first strand cDNA synthesis prior to conventional PCR (not shown) and RT-qPCRs.

As shown in Figure 6C, an induction of steady state mRNA levels was observed for *EucaDof3* in seedlings sprayed with water when compared to seedlings immediately frozen in liquid nitrogen (control in Figure 6). We assumed that such an effect was due to the mechanical stress imposed to plants. Similar induction or effect was also described in many works [39-42]. *EucaDof13*, on the other hand, was the only gene whose expression was reduced by the water treatment. All remaining *Dof* genes did not statistically change their expressions in response to the water spray. We therefore considered the water sprayed seedlings (H_2O in Figure 6) as controls for further analysis of *Dof* gene expression when ABA, NAA, KIN or NaCl were considered, since all these treatments were also perfomed by the spray of the respective solutions. Since plants submitted to cold and drought stresses were not sprayed, the effects of

these signals were considered taking into account plants immediately frozen in liquid nitrogen (Control in Figure 6C).

In our analysis, four out of ten *E. grandis Dof* genes tested by RT-qPCR exhibited increased levels of steady-state mRNA when seedlings were treated with ABA (Figure 6B). *EucaDof21* and *22* showed the highest increase in expression values, with about four times the level of relative expression found in water-treated seedlings. The expression levels of *EucaDof2* and *12* were also significantly increased, being approximately three times higher than the water-treated control. *EucaDof3* showed no difference, while *EucaDof4*, *13* and *17* showed no detectable levels of steady state mRNA levels.

Three out of ten *E. grandis Dof* genes tested by RT-qPCR exhibited increased levels of steady-state mRNA when plantlets were treated with NAA (Figure 6A). The highest expression values were show by *EucaDof2, 10*, *13*, with about three times the level of expression found in water sprayed seedlings, and *EucaDof22*, again exhibiting four times higher levels of expression. The expression levels of *EucaDof3, 17* and *21* did not statistically differed from water-treated seedlings.

When spraying seedlings with KIN, four out of ten *E. grandis Dof* genes showed increased levels of steady-state mRNA, reaching an eight-fold induction for *EucaDof2*, *13* and *21* (Figure 6A). *EucaDof22* followed with two-fold inductions and the remaining genes did not statistically changed their expressions in comparison to water-treated controls.

A clear opposite effect was generally observed in the patterns of *E. grandis Dof* gene expression when abiotic stresses were imposed on seedlings. When a sodium chloride solution (NaCl in Figure 6B) was sprayed on seedlings, all genes but *EucaDof2*, *13, 17* and *22* showed a strong decrease in expression in comparison to water-treated seedlings. Steady-state mRNA levels were ten times lower for *EucaDof21*, and down to four times for *EucaDof3*. *EucaDof2*, *13, 17* and *22* did not statistically change their expressions after NaCl treatment.

Remembering that seedlings submitted to drought and cold stresses were not sprayed and that *Dof* expression should be considered in respect to seedlings frozen in liquid nitrogen (Control in Figure 6C), *E. grandis Dof* genes exhibit a diverse pattern of expression in response to drought whereas the cold treatment determined a general decrease in expression, similar to that observed by NaCl (Figure 6B). While *EucaDof4*, *10* and *21* slightly increased their expressions in response to drought, *EucaDof12* and *13* showed four and two times reduction, with no statistical difference in the expressions of the remaining genes. Cold treatment, on the other hand, determined lower steady-state mRNA levels for all genes but *EucaDof3*, *4* and *21*. Surprisingly, *EucaDof21* did increase almost 2.5 times its expression following the cold treatment.

The expression profiles of some of the *A. thaliana Dof* genes were again profiled with the Genevestigator visualization tool [208] according to their available patterns of expression generated by microarrays. Results of this analysis in respect to gene responses to hormonal and stress signals here studied are shown in Figure 6A. We could not find any correlation between the patterns of expression of the *E. grandis Dof* genes tested and the available expression data of *A. thaliana Dof* genes. Considering the *E. grandis Dof* genes evaluated in our RT-qPCR assays and the available microarray data for *A. thaliana Dof* genes, only *EucaDof21* and *At1g21340.1* could be compared since they seem to be perfect orthologs (Figure 4). While *EucaDof21* showed clear increased expression in response to ABA, KIN, drought and cold (Figure 5B and 6C), for instance, At1g21340.1 showed no clear induction in response to the same signals (Figure 6A).

EucaDof4 encodes a peptide most similar to *A. thaliana* DAG2 protein (At2g46590.1, At2g46590.2; subgroup C2.1 in Figure 4 and Table 1). The DAG2-encoding gene was shown to exert effects on the control of seed germination [178, 179, 197]. The induction of *EucaDof4* expression in response to auxin (NAA) and drought and the reduction of its expression imposed by ABA, a hormone known to be important for the maintenance of seed dormancy and inhibition of seed germination, are in agreement with a possible role for *EucaDof4* in seed germination. Nevertheless, as already pointed out, few *Dof* genes were functionally characterized in other plant

systems and the lack of strong sequence homology to those that were characterized inpairs further speculation on the possible functions of *E. grandis Dof* genes.

Conclusion

Aiming the first description of genes encoding Dof transcription factors in a woody species, we identified 22 sequences in the genome of *E. grandis* as belonging to the Dof gene family. Comparison of genomic and transcriptomic sequences allowed us to predict gene structures, including intron/exon positions, and to determine amino acid sequences. Based on these deduced peptide sequences, a phylogenetic analysis was conducted with Dof sequences from *A. thaliana*. The expression patterns of ten selected *E. grandis Dof* genes were assessed in three different organs and in response to hormonal and abiotic stress signals. A quite diverse pattern of expression was exhibited by *E. grandis Dof* genes, revealing their involvements in various metabolisms.

Considering the extensive annotation work of genomes of several plant species including *A. thaliana* [221], *Populus trichocarpa* [222], *Malus x domestica* [223] and *Vitis vinifera* [224], the emergence of several databases containing detailed information about transcription factors found in several species of plants and algae, and the fact that these data have been used as the basis for selection of genes that comprise the family of Dof transcription factors in *E. grandis*, we assume that the totality of *Dof* sequences from *E. grandis* are represented in this study. Further elucidation of their roles in plant development and defense is under investigation.

Methods

Bioinformatics

The Basic Local Alignment Search Tool (BLAST) [65] was employed to mine the complete available genome (EUCAGEN Database) and transcriptome (GENOLYPTUS Database) of *E. grandis* for Dof transcription factor-encoding genes. As query, amino acid sequences derived from 38 plant species were employed, and these sequences were originally present at the Plant Transcription Factor Data Database (University of Potsdam, Germany) [118] and Plant Transcription Factor Database (Peking University, China) [119].

The scaffolds selected were subjected to a self-training algorithm to predict open reading frame (ORFs) in eukaryotic genomes, the GeneMark.hmm ES-3.0 [201]. Predicted sequences were confirmed by BlastP with sequences from The National Center for Biotechnology Information [65] and from The Arabidopsis Information Resource [221]. Protein domains were characterized with the help of the Pfam software [202]. The Artemis Release 12.0 software [203] was used to build up the architecture of *E. grandis Dof* genes and to determine their nucleotide and amino acid sequences.

Proteins motifs were determined by the Multiple Em for Motif Elicitation Multiple (MEME) 4.4.0 tool [205]. To evaluate the evolutionary relationships among Dof proteins, a combined phylogenetic analysis with 65 sequences from *A. thaliana* and 22 from *E. grandis* for a common tree, using the Molecular Evolutionary Genetics Analysis (MEGA) v.4.1 tool [206] and the result was visualized with the employment of the Figtree v1.3.1. software (http://tree.bio.ed.ac.uk/).

Plant Material

Field grown *E. grandis* trees (four-years-old) planted at the Hortoflorestal Barba Negra (Aracruz Celulose S.A.) in Barra do Ribeiro, RS, Brazil, were employed in

this study. Samples were collected in duplicate from two clonal, independent individuals. Flowers, leaves and xylem tissues were harvested, immediately frozen in liquid nitrogen and stored at -80 °C until RNA extraction.

E. grandis seeds were surface sterilized, germinated in glass flasks containing Murashige & Skoog basic medium and grown under controlled conditions (25 °C and photoperiod of 8 h light). Groups of two flasks containing 3 to 4 three-months-old seedlings were sprayed three times with the following solutions: 0.5 M abscisic acid (ABA); 2 mg/L 1-naphthaleneacetic acid (NAA); 0.2 mg/L kinetin (KIN); 0.5 M sodium chloride (NaCl); or distilled water (H_2O, water-sprayed control). After the sprays, flasks were closed and kept under culture conditions for an additional periodo of 16 h. After that, seedlings were removed, frozen in liquid nitrogen and stored at -80 ^{o}C. In order to simulate drought stress, two flasks were kept opened for 16 h until freezing. The cold treatment was done by exposing the flasks to 4 °C for 3 days and a recovery period of 24 h at 25 °C. Seedlings from two flasks were not submitted to any treatment, being frozen in liquid nitrogen to serve as controls.

RNA extraction and evaluation of gene expression by RT-qPCR

Total RNA was extracted from the duplicate samples of seedlings with EB-CTAB protocol [225] according to the small scale RNA isolation. Primer pairs for all gene sequences were designed using the PrimerQuest tool (http://www.idtdna.com/Scitools/Applications/Primerquest) and are listed in the Additional file 2. The relative transcript abundance was detected by SYBR Green and PCRs were carried out in a total volume of 20 µL using the StepOne™ *Real-Time* PCR System (Applied Biosystems). Reaction conditions were: one cycle of denaturation at 95 °C for 10 min followed by 40 cycles of 95 °C for 15 sec (denaturation) and 60 °C for 15 sec (annealing and elongation). The PCR was followed by a melting curve program (60-95 °C with a heating rate of 0.1 °C per second and a continuous fluorescence measurement). A negative control was run without cDNA template in all assays to assess the overall amplification specificity.

Authors' contributions

MCB carried out the bioinformatics analysis, PCR primer design, RT-qPCR analysis and drafted and edited the manuscript. GSDA performed the bioinformatics analysis. SSC participated in genomic and transcriptomic analysis, and the annotation of data. LFR helped in designing the study and in the bioinformatics analysis. JF and GP conceived and coordinated the study, drafted and revised the manuscript. GP gave the final approve. All authors read and approved the final paper.

Acknowledgements

The authors would like to acknowledge Dr. Charley Staats (Centro de Biotecnologia, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil) for the critical reading of the manuscript. We acknowledge M.Sc. Marta Dalpian Heis for helping with phylogenetic analysis. We are grateful to the staff of Aracruz Celulose S.A. for providing biological material. This work was financially supported by CNPq and CAPES, Brazil.

References

16. Rocha ME do N, Santos CL dos: **O uso comercial e popular do eucalipto *Eucalyptus globulus* Labill - Myrtaceae**. *Saúde e Ambiente* 2007, **2**:23-24.

33. Lorenzi H: *Árvores exóticas no Brasil: madeireiras, ornamentais e aromáticas*. Instituto Plantarum de Estudos da Flora; 2003.

38. Williams J, Woinarski J: *Eucalypt ecology individuals to eosystems*. Cambridge University Press; 1997, **11**:413.

39. Assis FT: **Melhoramento genético do eucalipto**. *Informe Agropecuário* 1986, **12**:36-46.

40. Hodgson LM: **Some aspects o flowering and reproductive behaviour in *Eucalyptus grandis* (Hill) Maiden at J.D.M. Keet Forest Reearch station**. *South African Forestry Journal* 1976, **98**:32-43.

41. Schumacher MV, Calil FN, Vogel HLMV: *Silvicultura Aplicada*. Santa Maria - RS: 2005:120.

42. Fobel M, Lynch DV, Thompson JE: **Membrane Deterioration in Senescing Carnation Flowers ' COORDINATED EFFECTS OF PHOSPHOLIPID DEGRADATION AND THE ACTION OF**. *Photosynthetica* 1987:204-211.

43. Panavas T, Rubinstein B: **Oxidative events during programmed cell death of daylily (*Hemerocallis hybrid*) petals**. *Plant Science* 1998, **133**:125 - 138.

65. Altschul SF, Madden TL, Schäffer a a, et al. **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucleic acids research* 1997, **25**:3389-402.

118. Perez-Rodriguez P, Riano-Pachon D, Correa LGG, et al. **PlnTFDB: updated content and new features of the plant transcription factor database**. *Nucleic Acids Research* 2009.

119. Guo AY, Chen X, Gao G, et al. **PlantTFDB: a comprehensive plant transcription factor database**. *Nucleic Acids Research* 2008, **36**:D966.

166. Yanagisawa S: **Dof domain proteins: plant-specific transcription factors associated with diverse phenomena unique to plants**. *Plant Cell Physiology* 2004, **45**:386-391.

169. Yanagisawa S: **The Dof family of plant transcription factors**. *Trends in Plant Science* 2002, **7**:555-560.

170. Yanagisawa S, Sheen J: **Involvement of maize Dof zinc finger proteins in tissue-specific and light-regulated gene expression**. *The Plant Cell* 1998, **10**:75-89.

171. Yanagisawa S: **Dof1 and Dof2 transcription factors are associated with expression of multiple genes involved in carbon metabolism in maize**. *Plant J.* 2000, **21**:281-288.

172. Papi M, Sabatini S, Altamura MM, et al. **Inactivation of the phloem-specific Dof zinc finger gene DAG1 affects response to light and integrity of the testa of *Arabidopsis* seeds**. *Plant Physiol.* 2002, **128**:411-417.

173. Park DH, Lim PO, Kim JS, et al. **The *Arabidopsis* COG1 gene encodes a Dof domain transcription factor and negatively regulates phytochrome signaling.** *The Plant Journal* 2003, **34**:161-71.

174. De Paolis A, Sabatini S, De Pascalis L, Costantino P, Capone I: **A rolB regulatory factor belongs to a new class of single zinc finger plant proteins**. *The Plant Journal* 1996, **10**:215-223.

175. Kisu Y, Ono T, Shimofurutani N, Suzuki M, Esaka M: **Characterization and expression of a new class of zinc finger protein that binds to silencer region of ascorbate oxidase gene.** *Plant & cell physiology* 1998, **39**:1054-64.

176. Mena M, Cejudo FJ, Isabel-Lamoneda I, Carbonero P: **A role for the DOF transcription factor BPBF in the regulation of gibberellin-responsive genes in barley aleurone**. *Plant Physiol* 2002, **130**:111-119.

177. Washio K: **Identication of Dof proteins with implication in the gibberellin-regulated expression of a peptidase gene following the germination of rice grains**. *Biochemica et Biophysica Acta* 2001, **1520**:54-62.

178. Gualberti G, Papi M, Bellucci L, et al. **Mutations in the Dof zinc finger genes DAG2 and DAG1 influence with opposite effects the germination of *Arabidopsis* seeds**. *Plant Cell* 2002, **14**:1253-1263.

179. Papi M, Sabatini S, Bouchez D, et al. **Identification and disruption of an *Arabidopsis* zinc finger gene controlling seed germination**. *Genes & Development* 2000, **14**:28-33.

180. Mena M, Vicente-Carbajosa J, Schmidt RJ, Carbonero P: **An endosperm-specific DOF protein from barley, highly conserved in wheat, binds to and activates transcription from the prolamin-box of a native B-hordein promoter in barley endosperm**. *Plant J.* 1998, **16**:53-62.

181. Plesch G, Ehrhardt T, Mueller-roeber B: **Involvement of TAAAG elements suggests a role for Dof transcription factors in guard cell-speci®c gene expression**. *The Plant Journal* 2001, **28**:455-464.

182. Gan S, Amasino RM: **Making Sense of Senescence (Molecular Genetic Regulation and Manipulation of Leaf Senescence).** *Plant physiology* 1997, **113**:313-319.

183. Ling H: **Sequence analysis of GDSL lipase gene family in *Arabidopsis thaliana***. *Pak. J. Biol. Sci* 2008, **11**:763–767.

184. Kim WY, Hicks KA, Somers DE: **Independent roles for EARLY FLOWERING 3 and ZEITLUPE in the control of circadian timing, hypocotyl length, and flowering time**. *Plant physiology* 2005, **139**:1557.

185. Nakamura T, Song I-J, Fukuda T, et al. **Characterization of TrcMADS1 gene of *Trillium camtschatcense* (Trilliaceae) reveals functional evolution of the SOC1/TM3-like gene family**. *Journal of Plant Research* , **118**:229-234.

186. Ferrario S, Busscher J, Franken J, et al. **Ectopic Expression of the Petunia MADS Box Gene UNSHAVEN Accelerates Flowering and Confers Leaf-Like Characteristics to Floral Organs in a Dominant-Negative Manner** . *The Plant Cell Online* 2004, **16** :1490-1505.

187. Lee H: **The AGAMOUS-LIKE 20 MADS domain protein integrates floral inductive pathways in *Arabidopsis***. *Genes & Development* 2000, **14**:2366-2376.

188. Lee S, Kim J, Han J-J, Han M-J, An G: **Functional analyses of the flowering time gene OsMADS50, the putative SUPPRESSOR OF OVEREXPRESSION OF CO 1/AGAMOUS-LIKE 20 (SOC1/AGL20) ortholog in rice**. *The Plant Journal* 2004, **38**:754-764.

189. Lee J, Oh M, Park H, Lee I: **SOC1 translocated to the nucleus by interaction with AGL24 directly regulates LEAFY**. *The Plant Journal* 2008, **55**:832-843.

190. Liu C, Zhou J, Bracha-Drori K, et al. **Specification of *Arabidopsis* floral meristem identity by repression of flowering time genes** . *Development* 2007, **134** :1901-1910.

191. Liu C, Xi W, Shen L, Tan C, Yu H: **Regulation of Floral Patterning by Flowering Time Genes**. *Developmental Cell* 2009, **16**:711-722.

192. Melzer S, Lens F, Gennen J, et al. **Flowering-time genes modulate meristem determinacy and growth form in *Arabidopsis thaliana***. *Nat Genet* 2008, **40**:1489-1492.

193. Lee J, Lee I: **Regulation and function of SOC1, a flowering pathway integrator.** *Journal of experimental botany* 2010, **61**:2247-2254.

194. Livak KJ, Schmittgen TD: **Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method.** *Methods (San Diego, Calif.)* 2001, **25**:402-8.

195. ABRAF: **Anuário estatístico da ABRAF 2010 ano base 2009**. *ABRAF - Brasília* 2010, **54**:140p.

196. Century K, Reuber TL, Ratcliffe OJ: **Regulating the regulators: the future prospects for transcription-factor-based agricultural biotechnology products.** *Plant physiology* 2008, **147**:20-9.

197. Lijavetzky D, Carbonero P, Vicente-Carbajosa J: **Genome-wide comparative phylogenetic analysis of the rice and *Arabidopsis* Dof gene families.** *BMC evolutionary biology* 2003, **3**:17.

198. Yanagisawa S: **A novel DNA-binding zinc finger motif that may form a single zinc finger motif**. *Nucleic Acids Research* 1995, **23**:3403-3410.

200. Thornton J: **Gene family evolution and homology: genomics meet phylogenetics Ann. Rev**. *Genomics Human Genet* 2000.

201. Lomsadze A, Ter-Hovhannisyan V, Chernoff YO, Borodovsky M: **Gene identification in novel eukaryotic genomes by self-training algorithm.** *Nucleic acids research* 2005, **33**:6494-506.

202. Bateman A, Birney E, Cerruti L, et al. **The Pfam protein families database**. *Nucleic acids research* 2002, **30**:276.

203. Carver T, Berriman M, Tivey A, et al. **Artemis and ACT: viewing, annotating and comparing sequences stored in a relational database.** *Bioinformatics (Oxford, England)* 2008, **24**:2672-6.

204. Guo A-Y, Chen X, Gao G, et al. **PlantTFDB: a comprehensive plant transcription factor database.** *Nucleic acids research* 2008, **36**:D966-9.

205. Bailey TL, Williams N, Misleh C, Li WW: **MEME: discovering and analyzing DNA and protein sequence motifs.** *Nucleic acids research* 2006, **34**:W369-73.

206. Tamura K, Dudley J, Nei M, Kumar S: **MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0**. *Molecular biology and evolution* 2007, **24**:1596.

207. Czechowski T, Stitt M, Altmann T, Udvardi MK: **Genome-Wide Identification and Testing of Superior Reference Genes for Transcript Normalization**. *Plant Physiology* 2005, **139**:5-17.

208. Hruz T, Laule O, Szabo G, et al. **Genevestigator v3: a reference expression database for the meta-analysis of transcriptomes.** *Advances in bioinformatics* 2008, **2008**:420747.

209. Czechowski T, Bari RP, Stitt M, Scheible W-R, Udvardi MK: **Real-time RT-PCR profiling of over 1400 *Arabidopsis* transcription factors: unprecedented sensitivity reveals novel root- and shoot-specific genes.** *The Plant journal : for cell and molecular biology* 2004, **38**:366-79.

210. Holland MJ: **Transcript abundance in yeast varies over six orders of magnitude.** *The Journal of biological chemistry* 2002, **277**:14363-6.

211. Imaizumi T, Schultz TF, Harmon FG, Ho LA, Kay SA: **FKF1 F-Box Protein Mediates Cyclic Degradation of a Repressor of CONSTANS in *Arabidopsis*** . *Science* 2005, **309** :293-297.

212. Kim H-S, Kim SJ, Abbasi N, et al. **The DOF transcription factor Dof5.1 influences leaf axial patterning by promoting Revoluta transcription in Arabidopsis**. *The Plant Journal* 2010, **64**:524-535.

213. Gardiner J, Sherr I, Scarpella E: **Expression of Dof gnene idnetitifies eraly stages of vascular development in *Arabidopsis* leaves**. *The International Journal of Developmental Biology* 2010.

214. Wellmer F, Alves-Ferreira M, Dubois A, Riechmann JL, Meyerowitz EM: **Genome-Wide Analysis of Gene Expression during Early *Arabidopsis* Flower Development**. *PLoS Genetics* 2006, **2**:e117.

215. Wei P-C, Tan F, Gao X-Q, et al. **Overexpression of AtDOF4.7, an Arabidopsis DOF Family Transcription Factor, Induces Floral Organ Abscission Deficiency in *Arabidopsis***. *PLANT PHYSIOLOGY* 2010, **153**:1031-1045.

216. Derbyshire P, Drea S, Shaw PJ, Doonan JH, Dolan L: **Proximal-distal patterns of transcription factor gene expression during *Arabidopsis* root development.** *Journal of experimental botany* 2008, **59**:235-45.

217. Braam J, Davis RW: **Rain-, wind-, and touch-induced expression of calmodulin and calmodulin-related genes in *Arabidopsis***. *Cell* 1990, **60**:357-364.

218. Braam J, Sistrunk ML, Polisensky DH, et al. **Life in a changing world: TCH gene regulation of expression and responses to environmental signals**. *Physiologia Plantarum* 1996, **98**:909-916.

219. Braam J, Sistrunk ML, Polisensky DH, et al. **Plant responses to environmental stress: regulation and functions of the *Arabidopsis* TCH genes.** *Planta* 1997, **203 Suppl**:S35-41.

220. Yi H-S, Heil M, Adame-Alvarez RM, Ballhorn DJ, Ryu C-M: **Airborne induction and priming of plant defenses against a bacterial pathogen** . *Plant Physiology*

221. Swarbreck D, Wilks C, Lamesch P, et al. **The Arabidopsis Information Resource (TAIR): gene structure and function annotation.** *Nucleic acids research* 2008, **36**:D1009-14.

222. others: **The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray)**. *Science* 2006, **313**:1596.

223. Velasco R, Zharkikh A, Affourtit J, et al. **The genome of the domesticated apple (Malus [times] domestica Borkh.)**. *Nat Genet* 2010, **42**:833-839.

224. Jaillon O, Aury J-M, Noel B, et al. **The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla**. *Nature* 2007, **449**:463-467.

225. Mamiatis T, Fritsch EF, Sambrook J, Engel J: *Molecular Cloning A Laboratory Manual*. 1985, **5**:104-104.

Figures

Figure 1 – Predicted gene architecture of 22 sequences of Dof transcription factors found in the *E. grandis* genome. The direction of the arrows indicates the direction of gene transcription. Introns are represented in red and exons are represented in blue. The location of the Dof conserved domain is shown by yellow arrows. Numbers represent the size in base pairs (bp) of each segment.

Figure 2 - Multiple sequence alignment of the 22 Dof domain sequences predicted from the *E. grandis* genome. Gene names correspond to those listed in Table 1. Cysteine residues putatively critical for the zinc-finger structures are marked in grey squares. Identical residues in all sequences are indicated by (*) under the column, conserved substitutions are indicated by (:), and semi-conserved substitutions are indicated by (.).

Figure 3 – Schematic representation of conserved motifs of *E. grandis* Dof proteins. Motifs were identified by means of the MEME software using the complete amino acid sequences of the 22 *E. grandis Dof* genes listed in Additional file 1. Positions of the identified motifs are relative to the Dof domain. Multilevel consensus sequences for the MEME defined motifs are listed in Table 2.

Figure 4 – Phylogenetic tree of *A. thaliana* and *E. grandis Dof*-deduced protein sequence families. The unrooted tree was inferred by the neighbour-joining method after the alignment of Dof amino acid sequences of 43 *A. thaliana* and 22 *E. grandis* Dof proteins. The topology of the unrooted tree allowed the identification of groups and subgroups of Major Clusters of Orthologous Genes (MCOGs) according to Lijavetzky *et al.*[197], shown in different hook colors. *E. grandis* sequences are represented in colors identical to MCOG groups and subgroups, while *A. thaliana* sequences are all represented in black. The scale bar corresponds to 0.5 estimated

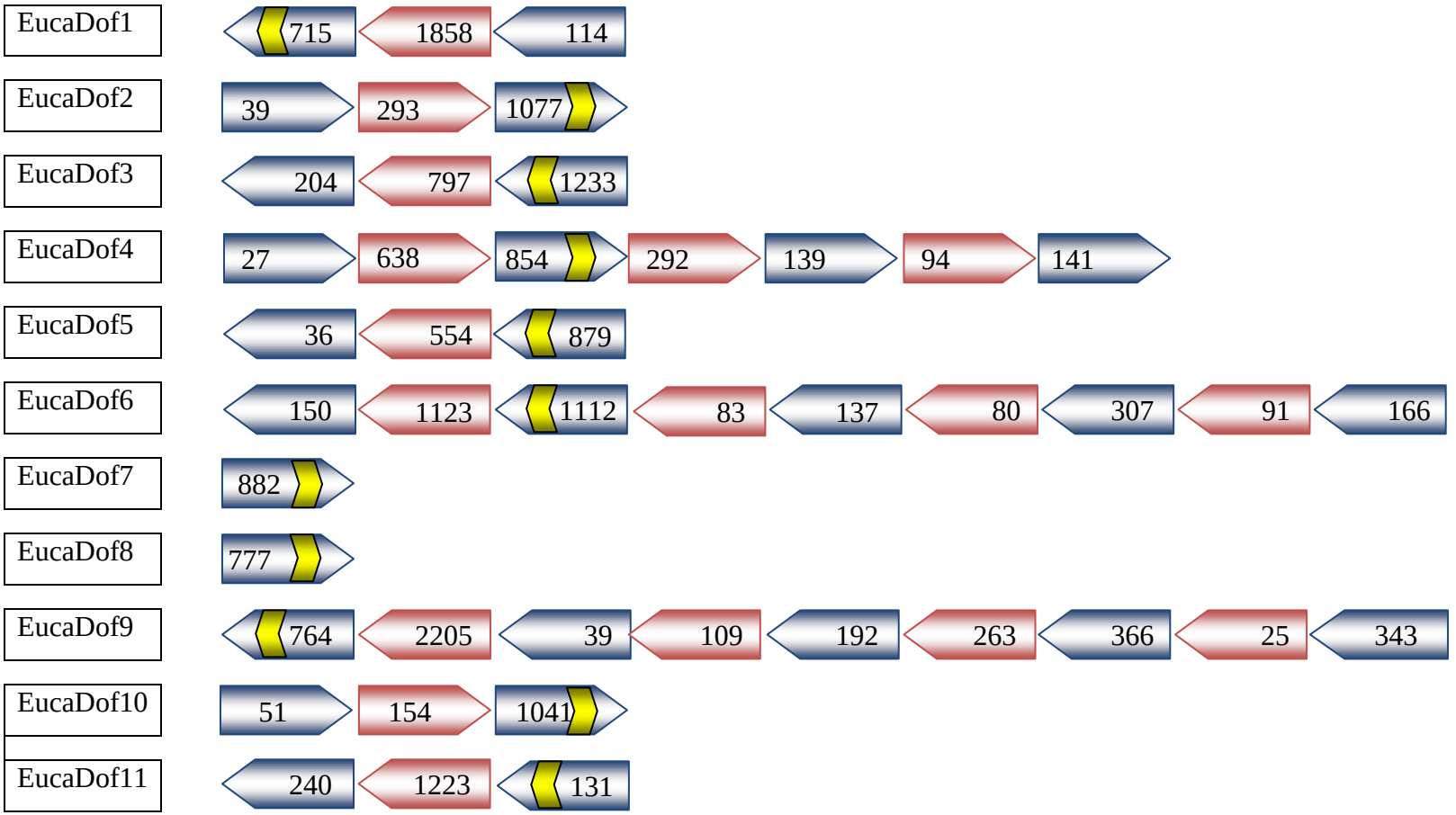
amino acid substitutions per site. The percentage above 70% of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) are represented by black dots below for the branches.

Figure 5 - **Relative spatial expression levels of selected *E. grandis Dof* genes by RT-qPCR.** (A) Microarray expression profiles of selected *A. thaliana Dof* genes generated by the Genevestigator tool based on data available in anatomy plant. Absolute values are rendered in blue-white, whereas the colors are normalized to the maximum value (the darkest color corresponds to the maximum value of an expression vector) – Absent $p>0.06$; Marginal $0.04 \leq p \leq 0.06$; Present $p<0.04$. (B and C) RT-qPCR relative spatial expression results of *E. grandis Dof* genes. Values found for flowers and xylem vascular tissues were normalized with values found in leaves. Gene expression was expressed in relation to that of the *RibL23A*, *Eucons08* and *Eucons21* constitutive, reference genes (Additional file 2). Error bars represent standard errors. Statistical analysis was done with a two-tailed, unpaired, Student´s t test, compared to leaves, $p<0.05$ (*).

Figure 6 –Relative expression levels of ten selected *E. grandis Dof* genes in response to different hormonal or stress signals by RT-qPCR. (A) Microarray expression profiles of selected *A. thaliana Dof* genes generated by the Genevestigator tool based on data available in response to external stimuli. Absolute values are rendered in green-black-red, whereas the colors are normalized to the maximum and minimum value – Minimum $p\geq-2.5$; Media $-0.5 \leq p \leq 0.5$; Maximum $p<2.5$. (B) RNA templates for cDNA and RT-qPCR were extracted from 90-days-old seedlings treated with water (H_2O) or treated with abscisic acid (ABA), 1-naphthaleneacetic acid (NAA), kinetin (KIN) or sodium chloride (NaCl) and (C) 90-days-old seedlings untreated (control) or subjected to cold or drought condition. Expression levels were normalized with untreated seedlings. Treatments whose expression was not detected are marked with (•) below the horizontal axis. Data were also normalized with values obtained for the *RibL23A*, *Eucons08* and *Eucons21* constitutive genes (Additional file 2). Results obtained are represented by a bar graphic, where error bars represent standard errors. Statistical

analysis was done with a two-tailed, unpaired, Student´s t test, compared to control and H_2O, $p<0.05$ (*).

Figure 1


EucaDof1
715
1858
114
EucaDof2
39
293
1077
EucaDof3
204
797
1233
EucaDof4
27
638
854
292
139
94
141
EucaDof5
36
554
879
EucaDof6
150
1123
1112
83
137
80
307
91
166
EucaDof7
882
EucaDof8
777
EucaDof9
764
2205
39
109
192
263
366
25
343
EucaDof10
51
154
1041
EucaDof11
240
1223
131

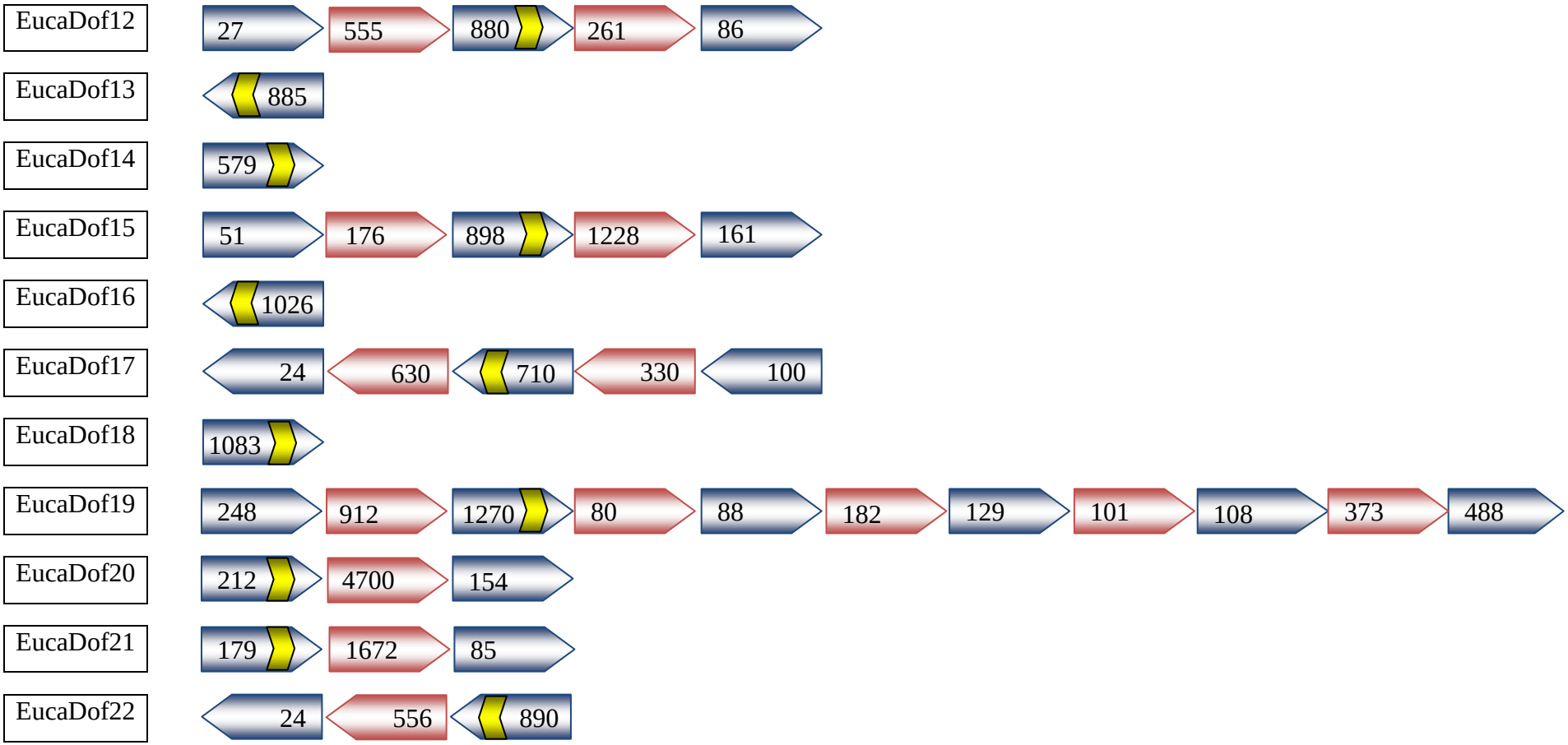

EucaDof12
27
555
880
261
86
EucaDof13
885
EucaDof14
579
EucaDof15
51
176
898
1228
161
EucaDof16
1026
EucaDof17
24
630
710
330
100
EucaDof18
1083
EucaDof19
248
912
1270
80
88
182
129
101
108
373
488
EucaDof20
212
4700
154
EucaDof21
179
1672
85
EucaDof22
24
556
890

Figure 2

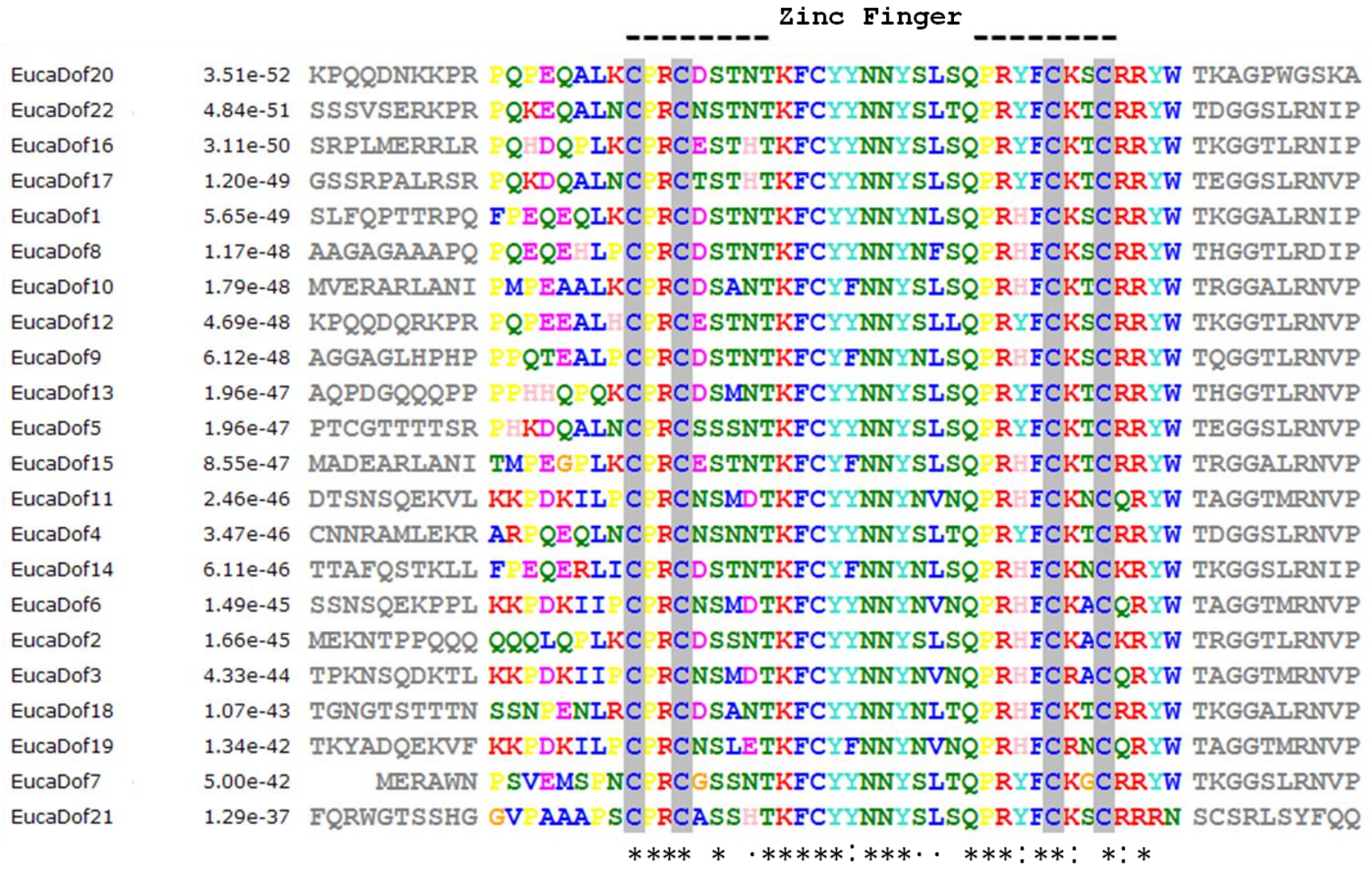

Zinc Finger
EucaDof20 3.51e-52 KPQQDNKKPR PQPEQALKCPRCDSTNTKFCYYNNYSLSQPRYFCKSCRRYW TKAGPWGSKA
EucaDof22 4.84e-51 SSSVSERKPR PQKEQALNCPRCNSTNTKFCYYNNYSLTQPRYFCKTCRRYW TDGGSLRNIP
EucaDof16 3.11e-50 SRPLMERRLR PQHDQPLKCPRCESTHTKFCYYNNYSLSQPRYFCKTCRRYW TKGGTLRNIP
EucaDof17 1.20e-49 GSSRPALRSR PQKDQALNCPRCTSTHTKFCYYNNYSLSQPRYFCKTCRRYW TEGGSLRNVP
EucaDof1 5.65e-49 SLFQPTTRPQ FPEQEQLKCPRCDSTNTKFCYYNNYNLSQPRHFCKSCRRYW TKGGALRNIP
EucaDof8 1.17e-48 AAGAGAAAPQ PQEQEHLPCPRCDSTNTKFCYYNNYNFSQPRHFCKSCRRYW THGGTLRDIP
EucaDof10 1.79e-48 MVERARLANI PMPEAALKCPRCDSANTKFCYFNNYSLSQPRHFCKTCRRYW TRGGALRNVP
EucaDof12 4.69e-48 KPQQDQRKPR PQPEEALHCPRCESTNTKFCYYNNYSLLQPRYFCKSCRRYW TKGGTLRNVP
EucaDof9 6.12e-48 AGGAGLHPHP PPQTEALPCPRCDSTNTKFCYFNNYNLSQPRHFCKSCRRYW TQGGTLRNVP
EucaDof13 1.96e-47 AQPDGQQQPP PPHHQPQKCPRCDSMNTKFCYYNNYSLSQPRYFCKTCRRYW THGGTLRNVP
EucaDof5 1.96e-47 PTCGTTTTSR PHKDQALNCPRCSSSNTKFCYYNNYSLSQPRYFCKTCRRYW TEGGSLRNVP
EucaDof15 8.55e-47 MADEARLANI TMPEGPLKCPRCESTNTKFCYFNNYSLSQPRHFCKTCRRYW TRGGALRNVP
EucaDof11 2.46e-46 DTSNSQEKVL KKPDKILPCPRCNSMDTKFCYYNNYNVNQPRHFCKNCQRYW TAGGTMRNVP
EucaDof4 3.47e-46 CNNRAMLEKR ARPQEQLNCPRCNSNNTKFCYYNNYSLTQPRYFCKTCRRYW TDGGSLRNVP
EucaDof14 6.11e-46 TTAFQSTKLL FPEQERLICPRCDSTNTKFCYFNNYNLSQPRHFCKNCKRYW TKGGSLRNIP
EucaDof6 1.49e-45 SSNSQEKPPL KKPDKIIPCPRCNSMDTKFCYYNNYNVNQPRHFCKACQRYW TAGGTMRNVP
EucaDof2 1.66e-45 MEKNTPPQQQ QQQLQPLKCPRCDSSNTKFCYYNNYSLSQPRHFCKACKRYW TRGGTLRNVP
EucaDof3 4.33e-44 TPKNSQDKTL KKPDKIIPCPRCNSMDTKFCYYNNYNVNQPRHFCRACQRYW TAGGTMRNVP
EucaDof18 1.07e-43 TGNGTSTTTN SSNPENLRCPRCDSANTKFCYYNNYNLTQPRHFCKTCRRYW TKGGALRNVP
EucaDof19 1.34e-42 TKYADQEKVF KKPDKILPCPRCNSLETKFCYFNNYNVNQPRHFCRNCQRYW TAGGTMRNVP
EucaDof7 5.00e-42 MERAWN PSVEMSPNCPRCGSSNTKFCYYNNYSLTQPRYFCKGCRRYW TKGGSLRNVP
EucaDof21 1.29e-37 FQRWGTSSHG GVPAAAPSCPRCASSHTKFCYYNNYSLSQPRYFCKSCRRRN SCSRLSYFQQ
**** * ·*****:***·· ***:**: *:*

Figure 3

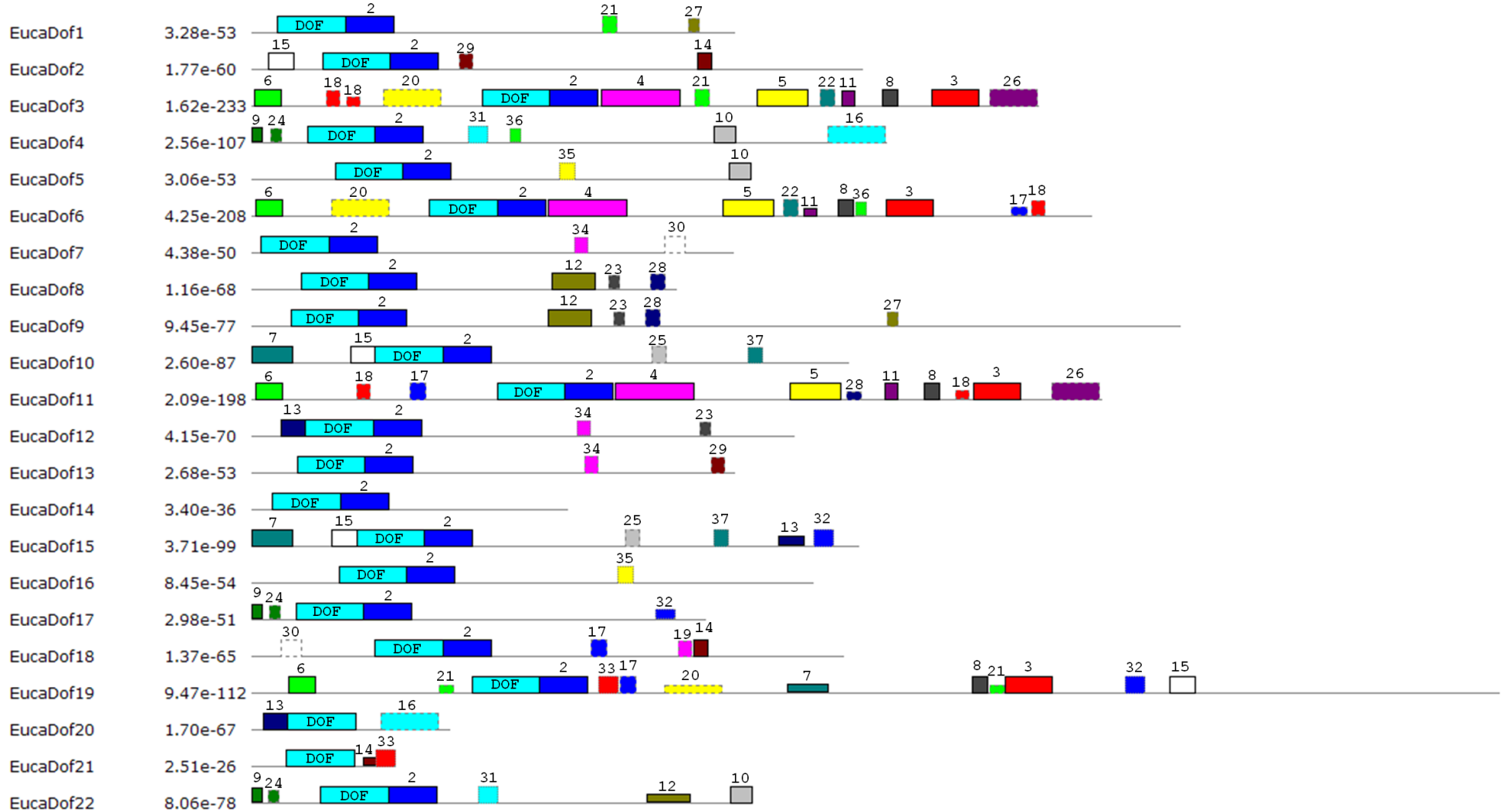

EucaDof1 3.28e-53
EucaDof2 1.77e-60
EucaDof3 1.62e-233
EucaDof4 2.56e-107
EucaDof5 3.06e-53
EucaDof6 4.25e-208
EucaDof7 4.38e-50
EucaDof8 1.16e-68
EucaDof9 9.45e-77
EucaDof10 2.60e-87
EucaDof11 2.09e-198
EucaDof12 4.15e-70
EucaDof13 2.68e-53
EucaDof14 3.40e-36
EucaDof15 3.71e-99
EucaDof16 8.45e-54
EucaDof17 2.98e-51
EucaDof18 1.37e-65
EucaDof19 9.47e-112
EucaDof20 1.70e-67
EucaDof21 2.51e-26
EucaDof22 8.06e-78
DOF

Figure 4

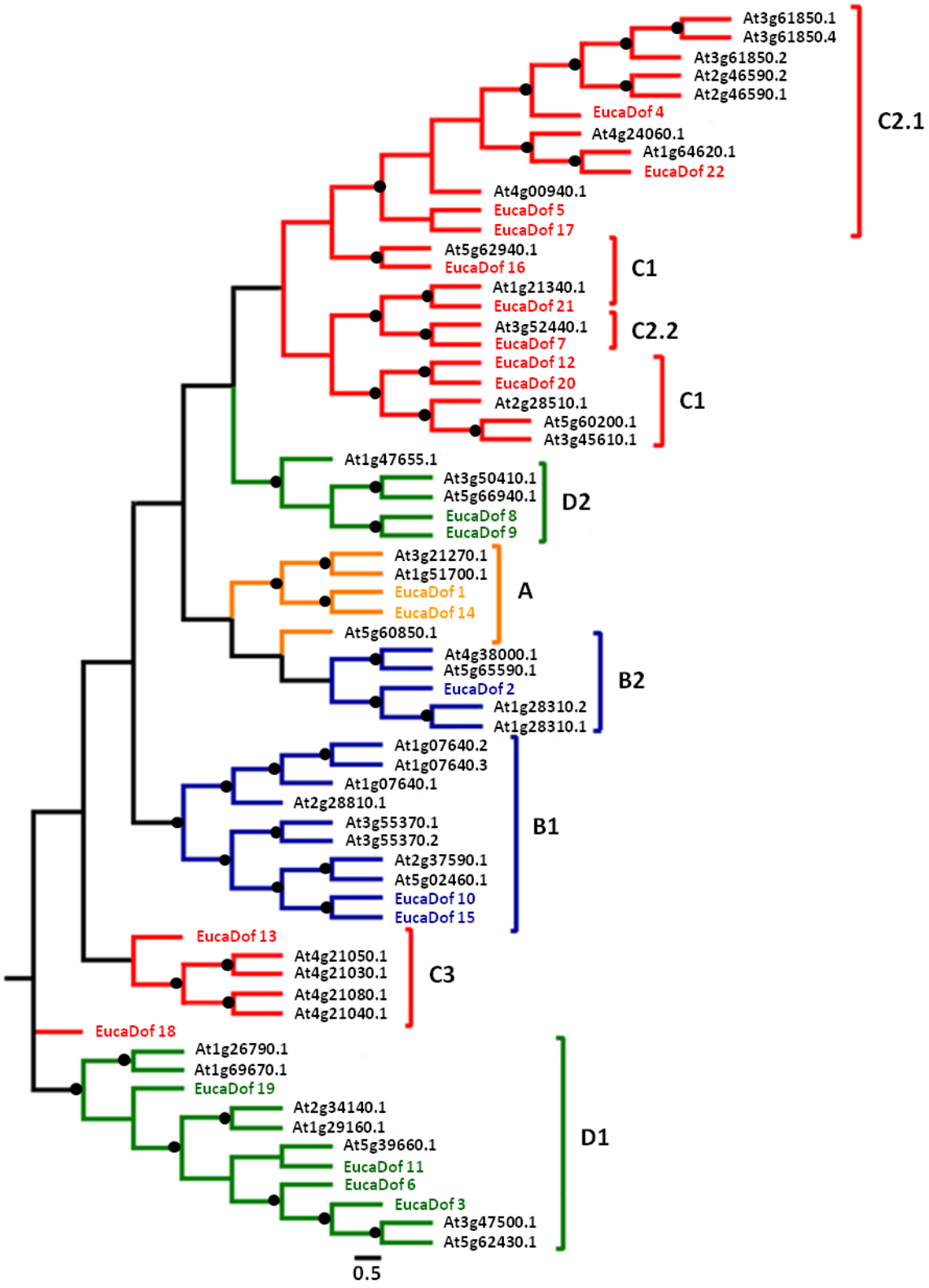
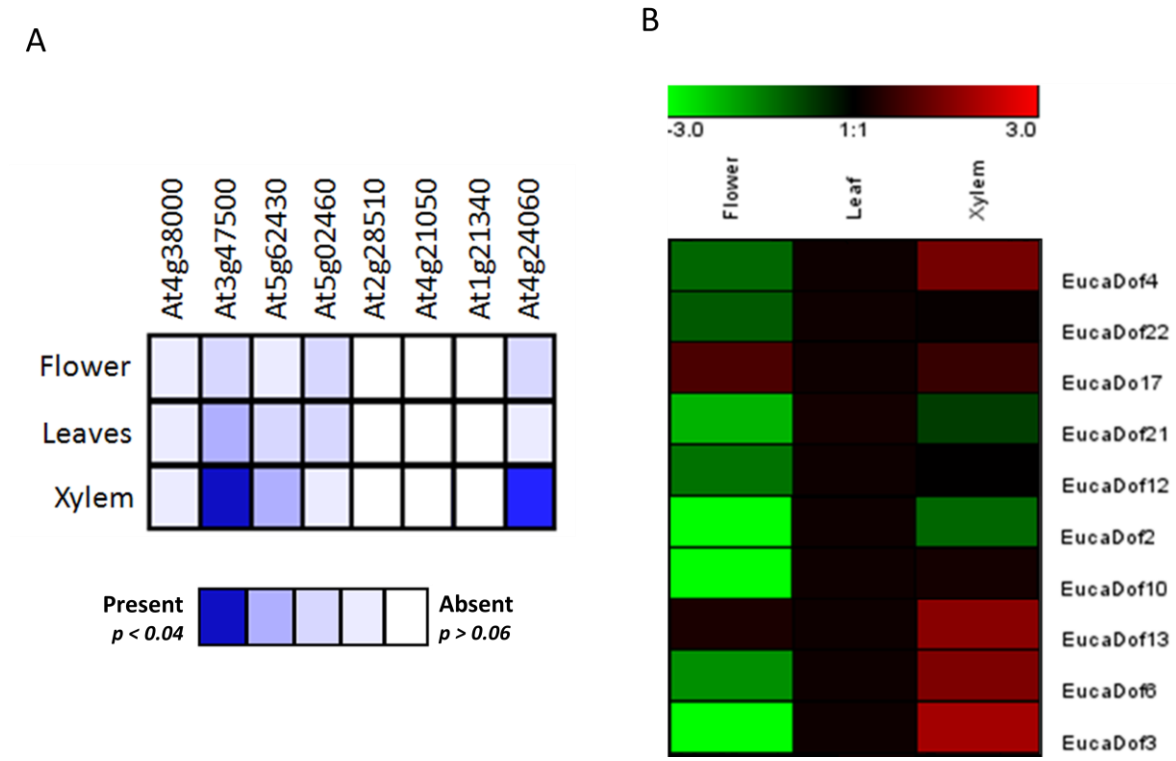

At3g61850.1
At3g61850.4
At3g61850.2
At2g46590.2
At2g46590.1
EucaDof 4
At4g24060.1
At1g64620.1
EucaDof 22
C2.1
At4g00940.1
EucaDof 5
EucaDof 17
At5g62940.1
EucaDof 16
C1
At1g21340.1
EucaDof 21
At3g52440.1
EucaDof 7
C2.2
EucaDof 12
EucaDof 20
At2g28510.1
At5g60200.1
At3g45610.1
C1
At1g47655.1
At3g50410.1
At5g66940.1
EucaDof 8
EucaDof 9
D2
At3g21270.1
At1g51700.1
EucaDof 1
EucaDof 14
A
At5g60850.1
At4g38000.1
At5g65590.1
EucaDof 2
At1g28310.2
At1g28310.1
B2
At1g07640.2
At1g07640.3
At1g07640.1
At2g28810.1
At3g55370.1
At3g55370.2
At2g37590.1
At5g02460.1
EucaDof 10
EucaDof 15
B1
EucaDof 13
At4g21050.1
At4g21030.1
At4g21080.1
At4g21040.1
C3
EucaDof 18
At1g26790.1
At1g69670.1
EucaDof 19
At2g34140.1
At1g29160.1
At5g39660.1
EucaDof 11
EucaDof 6
EucaDof 3
At3g47500.1
At5g62430.1
D1
0.5

Figure 5


A
At4g38000
At3g47500
At5g62430
At5g02460
At2g28510
At4g21050
At1g21340
At4g24060
Flower
Leaves
Xylem
Present
p < 0.04
Absent
p > 0.06
B
-3.0
1:1
3.0
Flower
Leaf
Xylem
EucaDof4
EucaDof22
EucaDo17
EucaDof21
EucaDof12
EucaDof2
EucaDof10
EucaDof13
EucaDof6
EucaDof3

C

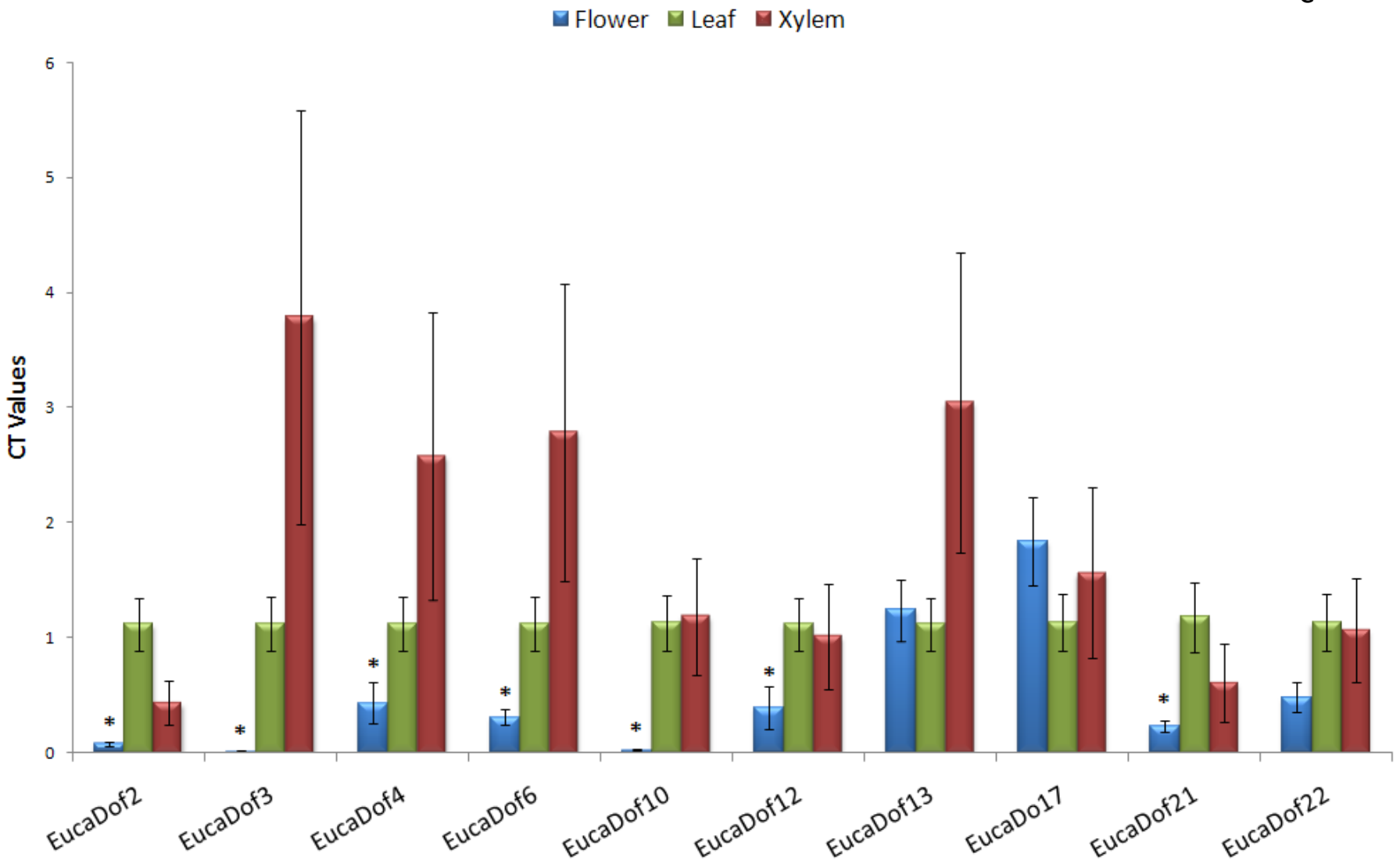
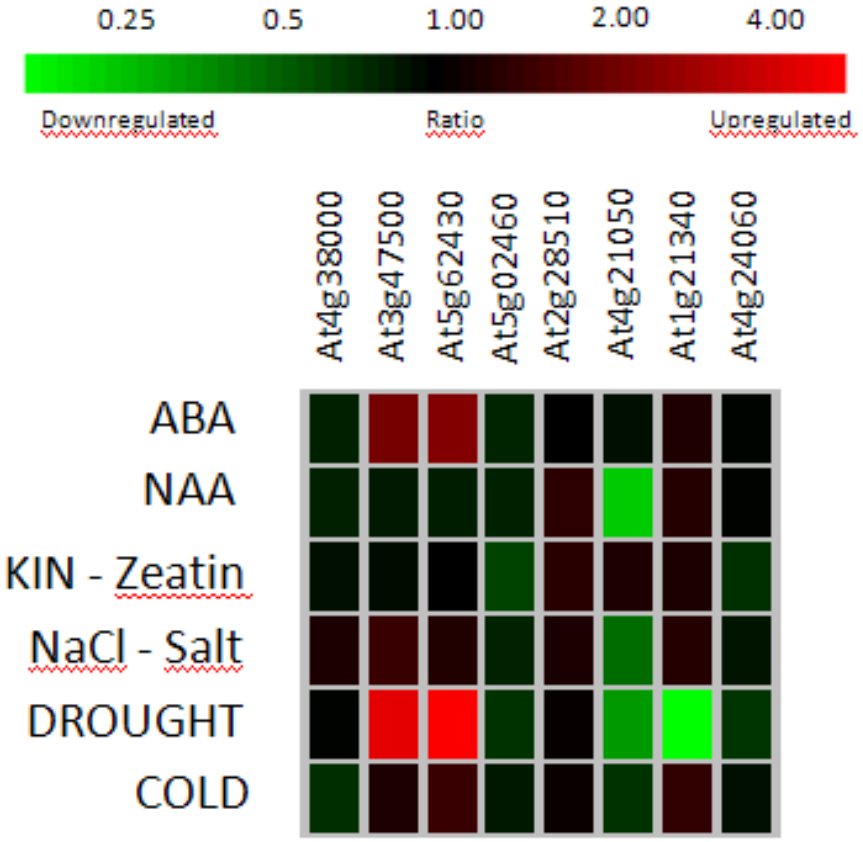
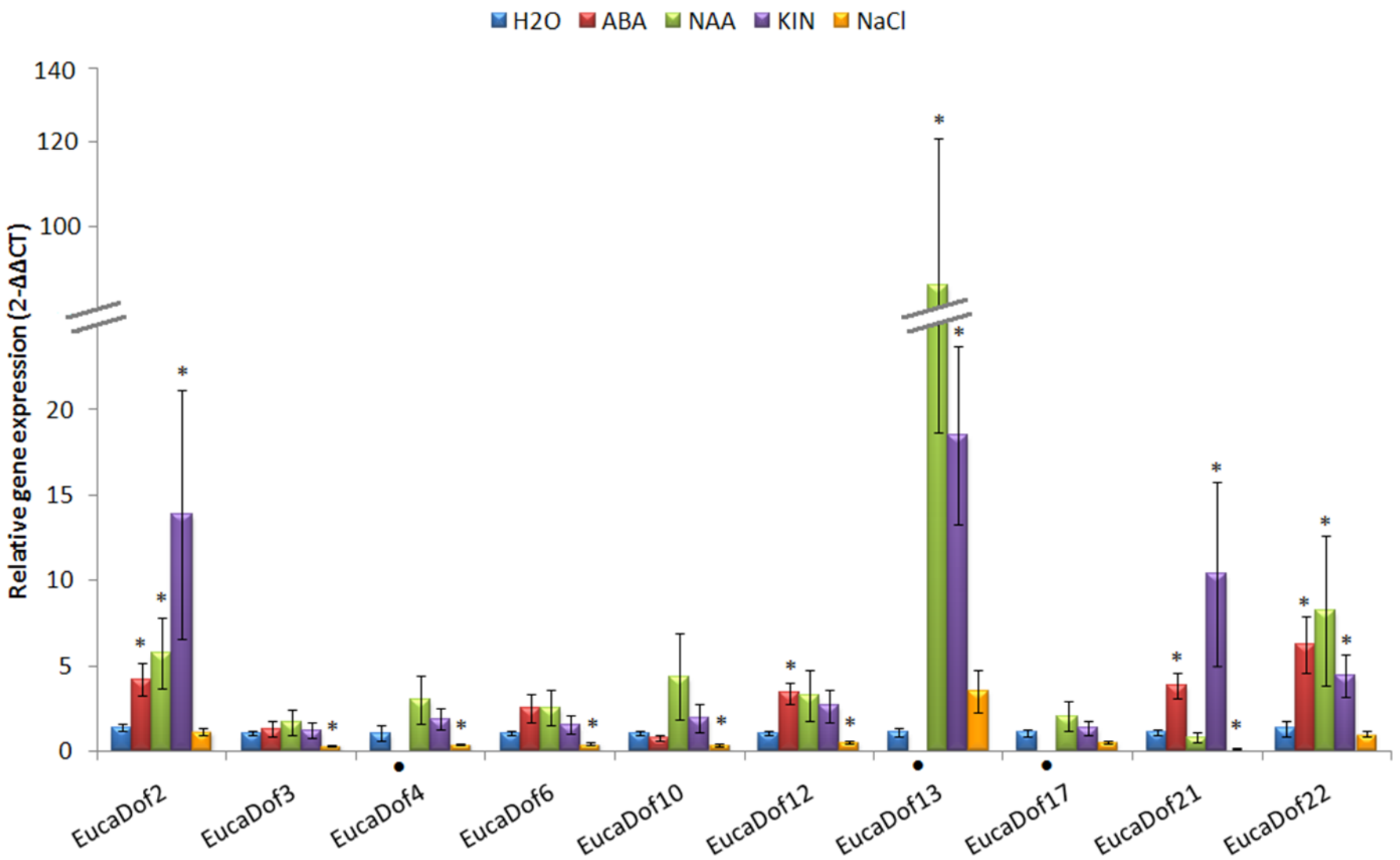

Flower
Leaf
Xylem
CT Values
6
5
4
3
2
1
0
*
EucaDof2
EucaDof3
EucaDof4
EucaDof6
EucaDof10
EucaDof12
EucaDof13
EucaDo17
EucaDof21
EucaDof22

Figure 6

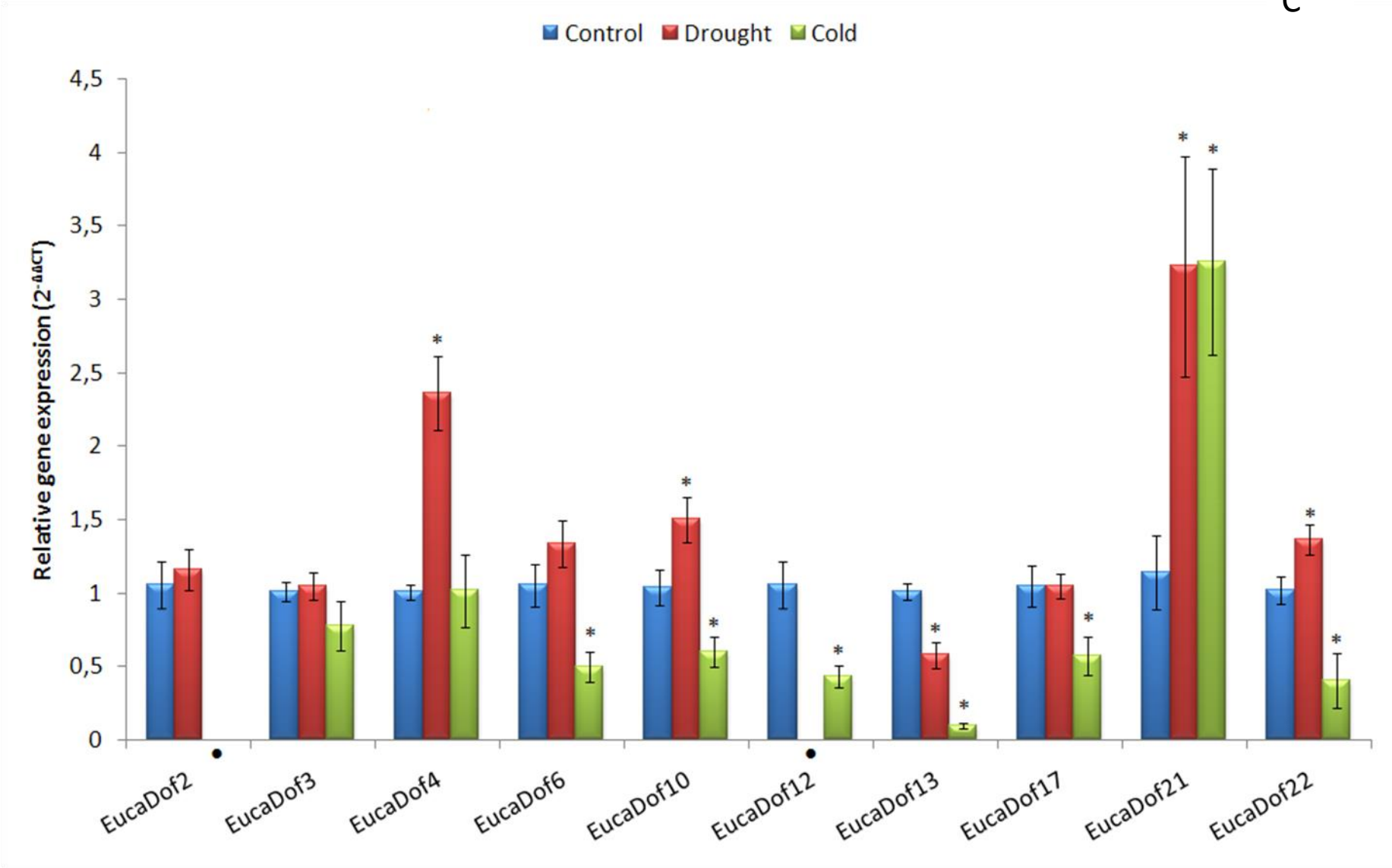
A


0.25
0.5
1.00
2.00
4.00
Downregulated
Ratio
Upregulated
At4g38000
At3g47500
At5g62430
At5g02460
At2g28510
At4g21050
At1g21340
At4g24060
ABA
NAA
KIN - Zeatin
NaCl - Salt
DROUGHT
COLD

B


H2O
ABA
NAA
KIN
NaCl
Relative gene expression (2-ΔΔCT)
140
120
100
20
15
10
5
0
EucaDof2
EucaDof3
EucaDof4
EucaDof6
EucaDof10
EucaDof12
EucaDof13
EucaDof17
EucaDof21
EucaDof22

C


Control
Drought
Cold
Relative gene expression (2-ΔΔCT)
4,5
4
3,5
3
2,5
2
1,5
1
0,5
0
EucaDof2
EucaDof3
EucaDof4
EucaDof6
EucaDof10
EucaDof12
EucaDof13
EucaDof17
EucaDof21
EucaDof22

Tables

Table 1 – *Eucalyptus grandis* Dof transcription factors annotation.

Table 2 – Group specific conserved motifs in *E. grandis*. Numbers correspond to the motif described in Figure 3. Sequences obtained from the analysis of the 22 Eucalyptus Dof complete proteins with the MEME system. Dof motif sequence in italics and bold. AT – *Arabidopsis thaliana*; EG – *Eucalyptus grandis*.

Table 1

Name	Size	Scaffold EUCAGEN	TAIR			NCBI		
			ID	Description	e-value	ID	Description	e-value
EucaDof1	293	4182	AT1G51700.1	ADOF1; DNA binding / transcription factor	2E-27	XP_002327532.1	f-box family protein [Populus trichocarpa]	1E-48
EucaDof2	371	978	AT1G28310.1	Dof-type zinc finger domain-containing protein	1E-26	XP_002274028.1	PREDICTED: hypothetical protein [Vitis vinifera]	2E-63
EucaDof3	478	735	AT3G47500.1	CDF3 (CYCLING DOF FACTOR 3); DNA binding / transcription factor	6E-72	gb|ACU80550.1	Dof1 protein [Jatropha curcas]	2E-125
EucaDof4	386	661	AT2G46590.2	DAG2 (DOF AFFECTING GERMINATION 2); DNA binding / transcription factor	1E-43	XP_002320208.1	predicted protein [Populus trichocarpa]	6E-72
EucaDof5	304	565	AT4G24060.1	Dof-type zinc finger domain-containing protein	2E-26	emb|CBI23042.3	unnamed protein product [Vitis vinifera]	3E-52
EucaDof6	514	531	AT3G47500.1	CDF3 (CYCLING DOF FACTOR 3); DNA binding / transcription factor	2E-66	gb|ACU80550.1	Dof1 protein [Jatropha curcas]	1E-113
EucaDof7	293	439	AT3G52440.1	Dof-type zinc finger domain-containing protein	4E-32	XP_002309254.1	f-box family protein [Populus trichocarpa]	9E-50
EucaDof8	258	356	AT3G50410.1	OBP1 (OBF BINDING PROTEIN 1); DNA binding / transcription factor	3E-34	XP_002310762.1	predicted protein [Populus trichocarpa]	8E-69

Name	Size	Scaffold EUCAGEN	TAIR			NCBI		
			ID	Description	e-value	ID	Description	e-value
EucaDof9	567	302	AT3G45610.1	Dof-type zinc finger domain-containing protein	3E-22	XP_002310762.1	predicted protein [Populus trichocarpa]	1E-48
EucaDof10	363	241	AT5G02460.1	Dof-type zinc finger domain-containing protein	2E-31	emb|CBI30569.3	unnamed protein product [Vitis vinifera]	3E-77
EucaDof11	516	220	AT5G39660.2	CDF2 (CYCLING DOF FACTOR 2); DNA binding / transcription factor	2E-79	XP_002283706.1	PREDICTED: hypothetical protein [Vitis vinifera]	5E-178
EucaDof12	330	213	AT2G28510.1	Dof-type zinc finger domain-containing protein	4E-32	emb|CBI16255.3	unnamed protein product [Vitis vinifera]	8E-57
EucaDof13	294	177	AT5G62940.1	Dof-type zinc finger domain-containing protein	4E-23	XP_002305784.1	f-box family protein [Populus trichocarpa]	8E-40
EucaDof14	192	61	AT1G51700.1	ADOF1; DNA binding / transcription factor	2E-27	XP_002327532.1	f-box family protein [Populus trichocarpa]	4E-36
EucaDof15	369	11	AT2G37590.1	ATDOF2.4 | Dof-type zinc finger domain-containing protein	1E-34	emb|CBI30569.3	unnamed protein product [Vitis vinifera]	8E-66

Name	Size	Scaffold EUCAGEN	TAIR			NCBI		
			ID	Description	e-value	ID	Description	e-value
EucaDof16	341	128	AT5G62940.1	Dof-type zinc finger domain-containing protein	1E-32	XP_002321600.1	f-box family protein [Populus trichocarpa]	1E-88
EucaDof17	277	4	AT4G24060.1	Dof-type zinc finger domain-containing protein	4E-28	XP_002275610.1	PREDICTED: hypothetical protein [Vitis vinifera]	2E-51
EucaDof18	360	689	AT5G60850.1	OBP4; DNA binding / transcription factor	8E-25	XP_002310753.1	f-box family protein [Populus trichocarpa]	1E-85
EucaDof19	776	1122	AT5G39660.2	CDF2 (CYCLING DOF FACTOR 2); DNA binding / transcription factor	3E-34	XP_002319195.1	predicted protein [Populus trichocarpa]	9E-109
EucaDof20	121	25	AT5G60200.1	Dof-type zinc finger domain-containing protein	3E-21	gb|EEF30118.1	zinc finger protein, putative [Ricinus communis]	8E-27
EucaDof21	87	84	AT5G62940.1	Dof-type zinc finger domain-containing protein	3E-09	XP_002890437.1	Dof-type zinc finger domain-containing protein [Arabidopsis lyrata]	1E-12
EucaDof22	304	152	AT4G24060.1	Dof-type zinc finger domain-containing protein	2E-45	XP_002518880.1	zinc finger protein, putative [Ricinus communis]	2E-93

Table 2

Motif	Domain Sequence	e-value	Species
1	***PPP[QDE]E[AP]L[KNP]CPRC[DN]S[TS]NTKFCY[YF]NNY[SN]L[ST]QPR[HY]FCK[GST]CRRYW***	8.9e-23	AT / EG
2	TKGG[TSA]LRN[VI]PVGGG[CS]RKNK[RK]S[SK]SSSSSSS	2.1e-87	AT / EG
3	[QR][KP][NE][GR][CS][VIL][LW]VPKTKRIDDP[SN]EAAKSSIW[AT]TLGIK[NK][ED][AE]	6.6e-11	AT / EG
4	[RQ]TN[HS]QFPF[SL]PTL[YQ]NLTQLGGIGLNLAA[TI]NGNN[QG][AG][HN][QG][AG][HN][QG][IN][GT]SS[LS][MF][ML][SN]DLGF[LF]H	5.1e-06	AT
5	[FT][FHM][LPS][PSY]G[QS][MI][MP][DE][SF]N[ST][VN]L[YP][SIT][SL][LPS][GP][FL][PQ][TS][ML][PGV]DY[KN][QPS]SN	2.7e-06	AT
6	GN[IT][HS][EG][NP][NV][NM][NG]NNNENNLM[AT]S[VL]GS[LS]S[PH]FALFD[PR]T[TM]GLY[NA]F[QP]N[DE][GV]N[IM]G	1.5e-05	AT
7	TSMVDSRVYQTPPVLMEEQPNLANLSRPVSGLTSPFNQTNQYFWPGSDF	9.1e-05	AT
8	RDP[AG]IKLFGK[TK]I[PT][FL]	2.9e-05	AT / EG
9	[KD][DLNLLSFPVMQD[HQ]HHH[AHG][LVM][EHS][LMH]	5.2e-04	AT
10	FSIDHHQGIGHNTINSNSNQRAQDNNDDMNGASRVLFPFSDM	2.0e-04	AT
11	[PI][MP][EM][QN][MN][IT][SH][SQ][TP]N[NT][TS][PN][QA][QR][QP][PN]T[FI][IL][AT][TS]N[TG][RV][PS][NT]A[TG]A[ST][NV][GS]G[SV][GS][GN]N[TN]NNTA[TV][MV][EA][TE]RK	1.5e-05	AT

Motif	Domain Sequence	e-value	Species
12	[PL][NS][GHPS][HP][GHP][GTV][AS][GQT][AG][GP][GLS][GSP][SQ][IAR][RM][PGV][GN]SM[AV][ED]RAR[LQI]A[NK][IV]	3.2e-04	AT / EG
13	ET[PL]VLQANPAA[LM]SRSMNF[HR]E	1.3e-04	AT / EG
14	[EH][LG][SN][TN][TN][QN][EM][KN]S[HS][GP]NN[TG]YW[NS]GMFS[NT]TGG[SG]S[WS]	9.7e-03	AT / EG
15	[MN][DH][RC]L[AD]F[GH][DEN][EG]SF[EQ]Q[DG][YL]YDVGS[DN][DN]L[IL][VDG]N[PQ][IL]	8.2e-02	AT
16	[MV]D[TAQ][AT][KQ]W[PTN]Q[GES][FI][QVN]	1.3e-02	AT / EG
17	DLNPPILFS[SN]QI[PH]NK	1.1e-02	AT
18	[CPV]FP[GP]VPM[PQ]Y[PTA]WN[PS][AT][FGM][PWY][PS]PP[FIP]Y	6.2e-03	AT / EG
19	[ST][AGNS][PS][NT]S[PS][TC]LGK[HR]SR[DE]E	6.6e-02	AT / EG
20	TDHQGLGHNSNNRSEALHSDHHQQGRVLFPFGDQMKELSSSITQEVDHDD	7.8e-02	AT
21	TLLQHPHKPMNNGGDMLGQSHLQTLASLQDLHVGGNNEDMKYKEGKLDQI	7.8e-02	AT
22	PHQISHFSSMNHHPLFYGLSDHMSSCNNNLPMIPSRFSDSSKTC	5.2e-01	AT
23	TGLDFGGTQISNMISGMSSSGGILDAERIPPSQQAQQFPFLINTGLVQ	6.8e-01	AT
24	[YMQ][LFI][DV][SPA][SQ][NS][WH]QQQ[QGH][NQ][QS][HNQ][HPQ]	9.0e-01	AT / EG
25	VSATQRNVKAEENDQDRGRDGVNNLSRNFLGNININSGRNEEYTSW	1.7e-01	AT
26	M[VI][SL[VA][EG][INT]Q[PQR][GI][NS][HN][QK][PS]F[FKRS][NKQ][VFT]QEN[NIV][DEHS][FLV][VS][GRV]S[FS]	1.9e-01	AT
27	H[HP][LH][QH]HQL[NQ]ENGS[IL][IV]SGH[GH][LQ]VLSH[HQ][LF]P	3.2e-01	AT

Motif	Domain Sequence	e-value	Species
28	LESEFLSSGFSSLSALGLGLPHQMSHDHTINGSFINNSTTNKPFLLSGLF	5.1e-01	AT
29	QFLQMPKMEGNGNITHQQQPSSSSSVYGSSSSPVALELLRTGVNV	8.2e-01	AT
30	[DG][PE][NST]R[VM][LF][WY]G[FC]P[WI][QG][MD][NQ]	1.2e-01	AT
31	EALQ[AS][AV]R[ITV]D[AFS]PNG[ATV][HQ][LH]P[HNP][LG][KR][IPT]NG[TA][VLTGG][LS]D[APS][PS][AVL][CG][DEN]S[MQ][AS][PSV][AIV][LP][NH][LI][AS][DEN]K	7.3e-01	EG
32	M[DN]N[LF]NV[FV][AR]NED	4.9e-01	AT
33	AA[AT]A[AS]HG[GS]FRHDFP[MV][KRLRCY][ST]DGQSC	2.7e-01	AT
34	INGFFMSSSSSLDPSNYNNMENNASVVNGAWLPLTNNNV	4.7e-01	AT
35	VKP[MIL][EN][EQD]I[VM]	1.5e-00	AT / EG
36	V[AGST][AI]V[GP]N[HLR]F[GDS][SDY]L[SY][EDH][IL]HG[GIV][MT][NV]	3.2e-00	AT
37	E[QS][IQ]DL[NA]L[AV][FY][AS]	1.1e-00	AT / EG

Additional Files

Additional file 1 - *Eucalyptus grandis* Dof genes, cDNAs and deduced protein sequences.

Additional file 2 - Sequence primers used in RT-qPCR. RibL23A, Eucons08 and Eucons21 (in italic) are reference, normalizing genes.

Additional file 1

>EucaDof1 _Eucalyptus_grandis_genomic sequence_reverse
ATGCAGCAGGACCCGTCATCGCTATTTCAGCCCACGACGAGGCCTCAATTCCCGGAGCAA
GAGCAGCTGAAATGCCCGCGCTGCGATTCCACCAACACCAAGTTCTGCTACTACAACAAC
TACAACCTCTCCCAGCCGCGCCACTTCTGCAAGAGCTGCCGCCGCTACTGGACCAAAGGC
GGCGCCCTCCGCAACATCCCCGTCGGCGGCGGCACCCGGAAGACCGCCAAGCGGGCCTCC
TCCAAGTCCAAGCGCCACAATGCCGGGTCTGGTTCGGACGCGGGCGCGGACCACCCGCAA
GCCCCTGGGTCGAGCCAGGCCCACAGCCGGACCACCGAAGAGGCGTCGCAATCTCCGCCG
CCAGGCTTCGTCGGGCCGACGTTGGATCTGGATCCGGCCCGGTTTTACGGCTCGCCATCC
GACCTGGAAAGGAGGATGCTTGATGTGAACGGATCCTTCAGCTCCCTCCTGGGGACGAGC
GGGCCATTCTTGAGCATCTTGGAGGGTCCGAATTTGAACGCCTCGGGAGCGAGCAGCATG
GTGCGGTTCGGGGGCCTCGGGGACGAAGTGAAACCGGGTCATGGCCACGTCCCGGGGTCG
GCTCGCGAGCCGGGCGCAGCGGCGGACGGGGCCGAGAATTTCATGGACATCCAGAACGGG
GATGCGGGTTGCTGGACAGGCGGGAGCGGATGGCCCGACTCGCCATCTATCCCCGGTTTG
AGTTTTCACGATAGAGAGAGATAGAGAGATTACTACTAAAATTTGGTCGATTTGTCTATT
CCCTCCCGTTTTCCTATCTCTTCCCTCATTTATCTTCACGGTGACGAGTTTCCTGGCATC
GGAATGTAGATGAGAAAGCGGAAGAAGGTACATTACACACTTTAGTATATACCGCTAAAA
CATATGATCATAGGTATGCTTTTGGTCTTGGATTATGTTGTAGCCCCTCTTGTGCTGGAT
CTATGTTGTCCCCACGACTTACTTATGTGACTGAATTTACAAACTACAGTAGGCTTTGTG
ATATAGGGTGTGCTTTGCACTTGTAGTGTTCGACTTTTTTACCAAGAAGAAGAAGAAGTG
ATGTTGAATCCATTGCGCAAGAGAGAGGAAGGAATGCAAGCATTTTGGCATTTTGCTTCT
GTGCTTTTTCTTTGGGAGTTCTGCAAGTCATTCTTACAATGTGGGGGGTGATGATTTGAG
ATGATGAGGGCATGTTCTGACTGAGATGATGAGGTGGGACTGGGAGTGGGTGAAAGATAT
GGGAAGTGGGGGGGCAACAGGAGTCTCTTGGTTTCTGGTTGAATGGATTTTGCGGTTTGT
CTTGTAAAGCATGGACAAAGAACTGGTTGTGATCTCATTGGATATTATTTTATTGCCGGA
AGTTTCTATCTTTGTCGCCTCTTTTTTATTTTCTGTAAAGTAGTGGCGAAATCGTCCAAC
ACTCCCTTTCACCTTCTAAAACGAGTCAAAATGCAAATTTGGTTTTTAGCAATGGAAGAA
GAAGGTGCTCTGGCATGGCTCTTCTTCTTGTTGGATGATAGCGTTTCCTCTGCACTATAT
GACGATATAGAGCATTGGCTCATGTGAAATTCCTCCTGAAAAACCTGTGGGAGGGACAGG
GATTTAATGACTAATCTTCCGTGAGCTCCAACGACTAATAAGATATATGACGAACTCCTC
GTTTCGGGTATACCTAATCGGTAAGGCCCGTGTAGGCTTGGGCGAGCCAAGATAATAACA
AGCCGGTGCTGAAAAGGAAATCGCTTTACTTAAGTTCGTCTATTTCACCTTTGAGCCCGT
TCAGTAATCTATCGATGTGGTAACTTGTTGCGTAACATTTTAGAAACTGTCCACTGGGTT
TCAAGTTGATGAGGGTTTAAAATGTGTACCATTGAATATAACTTCATTGCACTCTTCTGA
CAATTAAATGAATTAGGAGGAAGTTTGTACGTTAATGTGAAGCGACCACCTTCCTGAAGT
GGATGGTGGTCCTCGCGGCAAGAAATGAAAATTGAACGGTCGATTCGTGAAAGGCCGGAA
ATTAGTTGAGCATGCCTGGTGAAGGAGGTGGCTAGATACAGCTTGACTTTTTTAATTTCT
GGCGTGATTGCGTTCTATCTTACTTCAAGCTAATTAACTAGGATTTGTACACAATTTTCC
GTGAAAAATGAAGGAAGAGTTAGCGACTCTAACGAATTTAATTTTGCTCACGTCAATGCA
AAAAGAAAGGCGGGTTCTCTGTAGTTGATAACAGAGTATCTTCTTTACTCGATTTATCGC
TTAGTAATCCTTAAGGAAAAGGATAGGAAAAGTGAGACTCTTAAAACTTTGAGTTTGTGG
AGTGGCTGCTTTGAGTTGATTGGCATAAGTAAAGTTAATCCTTTGCATGGCTCATTCCTC
TTTGATGCGACGGTACCGTTGAAAGCACCACTTGTGATTTGAATCATCTCTTGTGATGAT
CAAGTACATTTCTTTTGTTAATACGAATTTTAAGCGAACGAATCAACAAACGACACATTA
ACCGACCCAAAAAAGAAGGAAAAAAATAGACCACACCTTTGTTTTTTTTAGGAGATTCAT
GCAAGACCTTGTCTTGTCTGAAGAACCCGATGGGCAATCCTTTGTCCCATCAAGGCGCGA
CGGTAAAGCCCCTTGGACAATCCGATGGAAATTCTCTAGGGGCATGA

>EucaDof1_Eucalyptus_grandis_CDS sequence_reverse
ATGCAGCAGGACCCGTCATCGCTATTTCAGCCCACGACGAGGCCTCAATTCCCGGAGCAA
GAGCAGCTGAAATGCCCGCGCTGCGATTCCACCAACACCAAGTTCTGCTACTACAACAAC
TACAACCTCTCCCAGCCGCGCCACTTCTGCAAGAGCTGCCGCCGCTACTGGACCAAAGGC
GGCGCCCTCCGCAACATCCCCGTCGGCGGCGGCACCCGGAAGACCGCCAAGCGGGCCTCC
TCCAAGTCCAAGCGCCACAATGCCGGGTCTGGTTCGGACGCGGGCGCGGACCACCCGCAA
GCCCCTGGGTCGAGCCAGGCCCACAGCCGGACCACCGAAGAGGCGTCGCAATCTCCGCCG
CCAGGCTTCGTCGGGCCGACGTTGGATCTGGATCCGGCCCGGTTTTACGGCTCGCCATCC

GACCTGGAAAGGAGGATGCTTGATGTGAACGGATCCTTCAGCTCCCTCCTGGGGACGAGC
GGGCCATTCTTGAGCATCTTGGAGGGTCCGAATTTGAACGCCTCGGGAGCGAGCAGCATG
GTGCGGTTCGGGGGCCTCGGGGACGAAGTGAAACCGGGTCATGGCCACGTCCCGGGGTCG
GCTCGCGAGCCGGGCGCAGCGGCGGACGGGGCCGAGAATTTCATGGACATCCAGAACGGG
GATGCGGGTTGCTGGACAGGCGGGAGCGGATGGCCCGACTCGCCATCTATCCCCGGATTC
ATGCAAGACCTTGTCTTGTCTGAAGAACCCGATGGGCAATCCTTTGTCCCATCAAGGCGC
GACGGTAAAGCCCCTTGGACAATCCGATGGAAATTCTCTAGGGGCATGATACTGA

>EucaDof1_Eucalyptus_grandis_protein sequence
MQQDPSSLFQPTTRPQFPEQEQLKCPRCDSTNTKFCYYNNYNLSQPRHFCKSCRRYWTKG
GALRNIPVGGGTRKTAKRASSKSKRHNAGSGSDAGADHPQAPGSSQAHSRTTEEASQSPP
PGFVGPTLDLDPARFYGSPSDLERRMLDVNGSFSSLLGTSGPFLSILEGPNLNASGASSM
VRFGGLGDEVKPGHGHVPGSAREPGAAADGAENFMDIQNGDAGCWTGGSGWPDSPSIPDS
CKTLSCLKNPMGNPLSHQGATVKPLGQSDGNSLGA*

>EucaDof2_Eucalyptus_grandis_genomic sequence_forward
ATGCTTGTCATCCCTTCAACCGCTAACGAATGGACACAGGTAACCCCCATTTCACTCTCA
AGTCTAATTTCCTTTCTTCCTATTCCTTCTTTCTTTTCTTTTTTTCTTTTCTTTTTGGGT
CTTGGCCAAGACACAACGCCTCACTAAGTTCACATTCAATTATTAGACAAACCAAAGCAT
GGCCCTCAAGTTTTTTGTTCCTTAGCTCTTGTCGCCGAATCTCTCTTCCTGTCTTGTTAT
TTTCCCACCCCCAAACACCCTTTGCTTTGTAACTTCTTGACCTCTTTTTTTGTTTTGCCT
CTCATTCCCTTCGTTGGCGTGTCCATTACCAGGATCAGATGATAGATGAGAAGCGCTTGA
ACAACAGTATGGCTTCAACAGGTAGTCCCCTCATGGAAAAAAACACTCCACCCCAACAAC
AGCAGCAGCAACTTCAACCTCTGAAATGTCCTAGATGTGATTCCTCCAACACCAAGTTCT
GCTACTACAACAACTACAGCTTGTCCCAGCCCAGGCACTTCTGCAAGGCCTGCAAGCGAT
ACTGGACCCGTGGTGGAACCCTAAGGAATGTCCCTGTCGGCGGTGGCTGCCGCAAGAACA
AGCGCGTCAAGCGCCCTGCCTCGGCCTCAGCCACCGACACGGCTCCTTCTTCTTCCGCTA
GTGCTACATCAGATCATCATCCCCCTCCTCATAAGACCCACATGGAGAATCTCTCCTCAA
GTTGCTCTAACCACAACAGCAACCCCATGAATTTGGGCCCTTTGTTTTATGGTAACCCAG
CTGATCAAATGAGCAATAATCTTGCATTTCCTAGGCTCACTTCTTCTTCTTCTCTGTATG
ATCTGCATCCTCAGTTGAGTGCTCTCGGGCTAGGGTTTTCGTCATCTGCGATCATCCCTG
CAAATTCTTGTGCAGACACCAGCAGTTTGCAAGAAGGGTTTAATGCTAATAATGCCTCAG
GCCTCAATGTTCAAGATTTGGCCCAAATCACGAACCCACTTCTTCTTCCAAGCTATAGTT
CCATGTTTGGCACTTCTTCATCACCATCATCTTCCTCTTCTTGTTCCACCACCACCACCA
TCGCTTCTCTGCTTGCATCTAGCTTTCACCATCAGCAGCAGAAATTCATCAATGCCTCAG
GCCTGAGAGAATCTAGAGCACTAGTCCATAATAATTTCAAGCCTATGTCTCCAGTGGAAG
AGTTTCAAGGGCTTCGTAGCAACAGTGAAGTCAGGATTGAAGGAGGCCAAAGCAGAAATA
TAGGGTGGAATGGGAGTATTCCGTGCCAGAACCCGGTGGATCAAATCGGGCTATCGGCCG
ATCCAGGATCGATTTGTTGGAGCTCATTAAATGTGGGCAATTCTTGGCATGATCCCACTA
GTATTGGTCTTCATCAGTCCCTTCTTTGA

>EucaDof2_Eucalyptus_grandis_CDS sequence_forward
ATGCTTGTCATCCCTTCAACCGCTAACGAATGGACACAGGATCAGATGATAGATGAGAAG
CGCTTGAACAACAGTATGGCTTCAACAGGTAGTCCCCTCATGGAAAAAAACACTCCACCC
CAACAACAGCAGCAGCAACTTCAACCTCTGAAATGTCCTAGATGTGATTCCTCCAACACC
AAGTTCTGCTACTACAACAACTACAGCTTGTCCCAGCCCAGGCACTTCTGCAAGGCCTGC
AAGCGATACTGGACCCGTGGTGGAACCCTAAGGAATGTCCCTGTCGGCGGTGGCTGCCGC
AAGAACAAGCGCGTCAAGCGCCCTGCCTCGGCCTCAGCCACCGACACGGCTCCTTCTTCT
TCCGCTAGTGCTACATCAGATCATCATCCCCCTCCTCATAAGACCCACATGGAGAATCTC
TCCTCAAGTTGCTCTAACCACAACAGCAACCCCATGAATTTGGGCCCTTTGTTTTATGGT
AACCCAGCTGATCAAATGAGCAATAATCTTGCATTTCCTAGGCTCACTTCTTCTTCTTCT
CTGTATGATCTGCATCCTCAGTTGAGTGCTCTCGGGCTAGGGTTTTCGTCATCTGCGATC
ATCCCTGCAAATTCTTGTGCAGACACCAGCAGTTTGCAAGAAGGGTTTAATGCTAATAAT
GCCTCAGGCCTCAATGTTCAAGATTTGGCCCAAATCACGAACCCACTTCTTCTTCCAAGC
TATAGTTCCATGTTTGGCACTTCTTCATCACCATCATCTTCCTCTTCTTGTTCCACCACC
ACCACCATCGCTTCTCTGCTTGCATCTAGCTTTCACCATCAGCAGCAGAAATTCATCAAT
GCCTCAGGCCTGAGAGAATCTAGAGCACTAGTCCATAATAATTTCAAGCCTATGTCTCCA
GTGGAAGAGTTTCAAGGGCTTCGTAGCAACAGTGAAGTCAGGATTGAAGGAGGCCAAAGC
AGAAATATAGGGTGGAATGGGAGTATTCCGTGCCAGAACCCGGTGGATCAAATCGGGCTA
TCGGCCGATCCAGGATCGATTTGTTGGAGCTCATTAAATGTGGGCAATTCTTGGCATGAT
CCCACTAGTATTGGTCTTCATCAGTCCCTTCTTTGA

>EucaDof2_Eucalypts_grandis_protein sequence
MLVIPSTANEWTQDQMIDEKRLNNSMASTGSPLMEKNTPPQQQQQQLQPLKCPRCDSSNT
KFCYYNNYSLSQPRHFCKACKRYWTRGGTLRNVPVGGGCRKNKRVKRPASASATDTAPSS
SASATSDHHPPPHKTHMENLSSSCSNHNSNPMNLGPLFYGNPADQMSNNLAFPRLTSSSS
LYDLHPQLSALGLGFSSSAIIPANSCADTSSLQEGFNANNASGLNVQDLAQITNPLLLPS
YSSMFGTSSSPSSSSSCSTTTTIASLLASSFHHQQQKFINASGLRESRALVHNNFKPMSP
VEEFQGLRSNSEVRIEGGQSRNIGWNGSIPCQNPVDQIGLSADPGSICWSSLNVGNSWHD
PTSIGLHQSLL*

>EucaDof3_Eucalyptus_grandis_genomic sequence_reverse
ATGCAAGAGACCCAGGACCCCGCCATCAAGCTCTTCGGTGCCAAGATCCCCGTCTCCTCC
GCCGCAGACGGCGCGGCCCTCGCGGCGGCGAAACCGGAGGAGAGGGAAGATGGAGGCGGA
GGCGGAGGAGGAGGAGGAGAGGAGGAGTACGATGAAGATGGAGGAGGAGGAGGAGAGGAA
CAAGAAATTGAAGATGCGGAGCAGAGGTATGCTTGGTGAACTGAAAAGGCGTTTTTTTGG
TGATTGCCCGTGTTTCTTACCGCTTTAAAGATTGATGGGTCGATTTTGCTTTTTTGCGTT
TTTCGCTTGGTTTCTTGCTTTGGATGCTGCTCTGTCTCTGGGAGTTCTCTGTTTCGGGTA
TCGTTGGTTCAGGAATACAGACAAATGTTGGTGGCCACTTTCTTGAATTTACTGAGTATT
TTCTCCATATACCGTGAAGATTTGAGCTTCTTTCTTTGGAAATATGGACTAGGGGAATTG
CGGATTGGGTTCAATACAGTTTGATTAGTTGCTTTTTCTCATGTGTGATGTCTTGTTGAT
TCAGTCTCTCTCTTTTTTTCCCCTTTTGACATGTCGTAAAGGCCTTTTCTTTTTTGGGGG
GTTTCGTGTGGGAATCAATTCTTCTGCAATCTGCTAGAGGTTTCTCTCCCAGCAGCTCAA
CTGATCTTTGTCATTGTCCCTTTGTTTTGGTCCTAAAAACTATGGTTTCTTGTGCAATCG
GGATACATCTAGACGTGAAGCTAGAGGGTCAATGAAGTTTGTAACTTTGTGATTCTAGAA
CCTTCTGAGTTTCAGATGATAGCATTATACCTTCAAGGCAAAGTTTCCATAGGGACCTAA
AATCAAAAGACAAGCTTTGAAAGTTGCAATTCTTTAGGCTGATTGCCCTTCCTTCTCCTG
CTGTCTGCATACTCGGTTTATATAGTAATCCACTGATCAATCTTACTACTTTCAATCCTT
TCTTTTCATATTAATCCTGATTATGTTTTCACATGTGAAAGTTTCAGAACGCTTCACCTG
AAAAGGCTTCTGAGAATATGCAAGAAGATTGCGAACCGCCTCCAAATTCTCAGGAACTGG
ATCCCTCCAGGACGTCTCCAGATGCTGTGAACCCTATGACACCCTCTATTCAAGAAGAGA
GTACGACATCACATATCACTGAGACTGGAGCACAAGAAGAGCATGGCACGCCCAAAAATT
CGCAAGACAAAACACTGAAGAAACCAGACAAGATCATTCCTTGCCCTCGATGTAACAGCA
TGGACACAAAATTCTGTTACTACAACAATTACAATGTCAATCAGCCCCGGCACTTCTGCA
GAGCCTGTCAAAGGTACTGGACTGCAGGGGGTACTATGAGGAATGTTCCTGTGGGAGCTG
GACGGCGGAAGAACAAGAACGCTGCCACACGTTATCGCCCCATCACCATCTCCGAGGCTC
TCCAAGCAGCCAGAGTTGATTCCCCAAATGGGGCTCAACTCCCCCATTTAAGAACCAATG
GAGCCGTCCTTACCTTTGGTTTGGACTCCCCTGTGTGCAATTCTCAGTCCCCCATTCCAA
ACATCTCCGAGAAGAAGAAAGTGTTGAATGGTTACAAAAATGGGTTTCGGGACATTGAAA
ATGGGGATGATAGCTCCACCTCTTCCTCGGTCACAATATCGAATTCCATGAATGAGAAAG
GCCTAAATTTCCCACAGGAACAAGGGATGGGAAATGTTAATGGATTTTCTACGCAGATAC
CTTGCCTTCCCGGCGTTCCTTGGCCTTATCCATGGAATTCCACAATTCCCTCACCTGCCT
TTTGCCCTTCAGGATGTCCACCGATGTCATTTTACCCTGCAGCCTTTTGGGCTTGTGGTG
TTCTTCCTAATGCCTGGAACATTCCTTGTATGTCACCTCAGTCACCCTTTTGCAACCAAA
AGGCCACAACTTCTGGTCCAAATGCACCGACACTGGGAAAGCATTCTAGAGAATCTGAGC
CACTTAAATCGGATATTTCGCTGGAACAAGCGGAACCCCCAAGACAGAAGAATGGTAGCA
TTTTGGTTCCGAAGACTCTGCGTATAGATGATCCCAGCGAGGCTGCCAAAAGTTCAATAT
GGGCGGCCCTCGGTATTAAGGCCTTCGAACCAAAGTCCAGAGAAGGCAAAGTTAAAGATG
CTGAGACCTTGCCACATTTGCATGCTAATCCTGCTGCATTGTCTCGTTCCCTCAATTTCC
ATGAGAACACTTGA

>EucaDof3_Eucalyptus_grandis_CDS sequence_reverse
ATGCAAGAGACCCAGGACCCCGCCATCAAGCTCTTCGGTGCCAAGATCCCCGTCTCCTCC
GCCGCAGACGGCGCGGCCCTCGCGGCGGCGAAACCGGAGGAGAGGGAAGATGGAGGCGGA
GGCGGAGGAGGAGGAGGAGAGGAGGAGTACGATGAAGATGGAGGAGGAGGAGGAGAGGAA
CAAGAAATTGAAGATGCGGAGCAGTTTCAGAACGCTTCACCTGAAAAGGCTTCTGAGAAT
ATGCAAGAAGATTGCGAACCGCCTCCAAATTCTCAGGAACTGGATCCCTCCAGGACGTCT
CCAGATGCTGTGAACCCTATGACACCCTCTATTCAAGAAGAGAGTACGACATCACATATC
ACTGAGACTGGAGCACAAGAAGAGCATGGCACGCCCAAAAATTCGCAAGACAAAACACTG
AAGAAACCAGACAAGATCATTCCTTGCCCTCGATGTAACAGCATGGACACAAAATTCTGT
TACTACAACAATTACAATGTCAATCAGCCCCGGCACTTCTGCAGAGCCTGTCAAAGGTAC
TGGACTGCAGGGGGTACTATGAGGAATGTTCCTGTGGGAGCTGGACGGCGGAAGAACAAG

AACGCTGCCACACGTTATCGCCCCATCACCATCTCCGAGGCTCTCCAAGCAGCCAGAGTT
GATTCCCCAAATGGGGCTCAACTCCCCCATTTAAGAACCAATGGAGCCGTCCTTACCTTT
GGTTTGGACTCCCCTGTGTGCAATTCTCAGTCCCCCATTCCAAACATCTCCGAGAAGAAG
AAAGTGTTGAATGGTTACAAAAATGGGTTTCGGGACATTGAAAATGGGGATGATAGCTCC
ACCTCTTCCTCGGTCACAATATCGAATTCCATGAATGAGAAAGGCCTAAATTTCCCACAG
GAACAAGGGATGGGAAATGTTAATGGATTTTCTACGCAGATACCTTGCCTTCCCGGCGTT
CCTTGGCCTTATCCATGGAATTCCACAATTCCCTCACCTGCCTTTTGCCCTTCAGGATGT
CCACCGATGTCATTTTACCCTGCAGCCTTTTGGGCTTGTGGTGTTCTTCCTAATGCCTGG
AACATTCCTTGTATGTCACCTCAGTCACCCTTTTGCAACCAAAAGGCCACAACTTCTGGT
CCAAATGCACCGACACTGGGAAAGCATTCTAGAGAATCTGAGCCACTTAAATCGGATATT
TCGCTGGAACAAGCGGAACCCCCAAGACAGAAGAATGGTAGCATTTTGGTTCCGAAGACT
CTGCGTATAGATGATCCCAGCGAGGCTGCCAAAAGTTCAATATGGGCGGCCCTCGGTATT
AAGGCCTTCGAACCAAAGTCCAGAGAAGGCAAAGTTAAAGATGCTGAGACCTTGCCACAT
TTGCATGCTAATCCTGCTGCATTGTCTCGTTCCCTCAATTTCCATGAGAACACTTGA

>EucaDof3_Eucalyptus_grandis_protein sequence
MQETQDPAIKLFGAKIPVSSAADGAALAAAKPEEREDGGGGGGGGEEEYDEDGGGGGEE
QEIEDAEQFQNASPEKASENMQEDCEPPPNSQELDPSRTSPDAVNPMTPSIQEESTTSHI
TETGAQEEHGTPKNSQDKTLKKPDKIIPCPRCNSMDTKFCYYNNYNVNQPRHFCRACQRY
WTAGGTMRNVPVGAGRRKNKNAATRYRPITISEALQAARVDSPNGAQLPHLRTNGAVLTF
GLDSPVCNSQSPIPNISEKKKVLNGYKNGFRDIENGDDSSTSSSVTISNSMNEKGLNFPQ
EQGMGNVNGFSTQIPCLPGVPWPYPWNSTIPSPAFCPSGCPPMSFYPAAFWACGVLPNAW
NIPCMSPQSPFCNQKATTSGPNAPTLGKHSRESEPLKSDISLEQAEPPRQKNGSILVPKT
LRIDDPSEAAKSSIWAALGIKAFEPKSREGKVKDAETLPHLHANPAALSRSLNFHENT*

>EucaDof4_Eucalyptus_grandis_genomic sequence_forward
ATGGATACTGCTCAATGGCCATCACAGGTAGAGACGTCCGTCTCTCTATTGTCCCTCTTG
CTTTCTTCTTTTTTTGCAATTCTCAGCTCTCCTTTTTTCCAGAGCTTTCTTTTTCCCTCT
CAGAGCTTCCATGGATTGAGAGGTTGTTTGTTGGGTGTCATTTTCTAACTGCACATAAGA
TGCATGGTGTTCTTTTGGCGTTGATCTTACTTTCATCATGTTGCTGTCCTCTCTTTTGCA
AAAGTGAAAACTTTAACTTTTTGAAATGTGAGGGCAATTGCTTGCTGCGACTTTTCCTTT
CTTCACCTTTTCTTTTGCTTATGTTAACTGTGATCATAAAGGAATAGGGGAAAAGGCATC
ATTTGTAATCAGCCTTTGTTCTTTTTATCTCTTCTCTCTCCCTTTCTTTCTCTTCAATTC
CTTTAACAGAAACTACTTCGGCCAGGGAGAAGAACAGTATCTAATATCTAGAAGCCCTAA
ACCGCTAGCTAATTGTCGATGTTCGATTTGATTCTCTCTAGCCTACATGCCTGAGGTGGT
TCTTATCACACAAGTCAGTGGTGAATTGTGAAGATTGATCTTTTACTCATAGTGGGAATA
AAAACCTTTCTCAACTCCCTCATCCTGTCAAACCCTAATTGTCGTGAGTCCTTACTGTGA
ACTAGGACACGAGGGTCGTCAAGGGAATGGAAGAGACATCTCCAAATGCATGCAATAATA
GGGCCATGTTGGAGAAGAGAGCAAGGCCTCAGGAGCAGCTGAATTGCCCAAGGTGTAATT
CGAACAACACCAAGTTCTGCTATTACAACAACTACAGCCTCACTCAACCCAGATACTTTT
GCAAGACTTGCAGAAGGTACTGGACCGATGGCGGGTCTCTAAGGAATGTCCCAGTGGGAG
GCGGTTCGAGGAAGAACAAGAAAGTATCCACCTCGTCAACGTCGTCGTCACCGTCAGCTT
CTTCTAGGCTCCCCGATCTTAACCCATCATCTGTAATCTCTCACTTCTCTGCTCAAAACC
CTAAGGCAGGACTACTGTACGAAGGTCAAGATCTCAATTTGGGCTTTCCAGGCACCCAAG
AATACCGCGCAACAATTGCTCAGTATGTTCAAGTTCCCAAAGTTGATCCTGAGACTAGTA
GTAGAGGTCCGGATAACAACTCTAGTGCACCTTCTTCAGCTCTAGAGCTGCTCCGGAATA
GTTCTAGTGGAATGGGTTCAAGGGGCTTGAGTCTTCCTTTCCTCCCAGCGAACATTCAGG
CCGATTCCAATGTTCTCTACTCATCGGCAGGCTTTAATTTGCAAGAACTCAAGCCCTCGC
TTGGCTTCTCTATGGATAGCAGGTATGCGAGTGCCCATGGAGTCGTTCATCAAGAGAATG
GTGGGAAGTTACTGTTCCCCTTTGGAGGAATGATGAAGCAGCTTTCGAGCACGAGTGGCA
GCGATCAAGTAGATCACAATGGAAAGGGTCAGAGCAGTTCGAATGCATATTGGAATGGAA
TGTTAGGTGGAGGTTCGTGGTAAAGAGGGATCGTTCAACTGGGGAAGCTGCGTGTGCTGG
CTCTTTTTCTCTCTTTTGCGTTAACCTTTTTTTTTCCTCTCATGGTTGTTACTGCATGTC
TGAAACTGTTCAGAGCTAACAGATAGAATACATGTATTTGTATGGCTTTCTTTTTTTTCC
TTCTAAACAATTTTAAGCTTTTTGGAGTCTTAGAAGGGTTTGCTTTTGCTAGCTACTCAA
GCCAAAAGTTGGATGAAGTTCATCGGTTTGGAGTTTACCTATCTGAGATGTGCTTTCTTA
TCATATTAAAGGAAGGAAAAGTTGGACTCAGAGGTCAATAGATACCTAAGGATTATTAGG
AGTGTCTTCCTGCCTGCTGAGTCTTTTGGTAAATACCCTGGACGATATCATGACGTTATT
AAAGGACGTTTCTCAAGTATCAATGAGCTGGTTTTCCTTAGCAAAGAACCTCGAACTCCG
ATCCATCTCGCACGTCTAAAGAAAAATTGTGAGATGGGCAGGGTTTTACGTAATTTCATT

ATAGAGCATGGTATCGAAGACTTTAATAATATCTGTGAATTTGTGTTTCTACGGCTTTGT
GGATTTCGATGTGCGGAAATTCTACTGTGTTGCTAAGCTGCTGTTGCATGTGCCTGTCAT
TGAGCTCATCGGTGAACCTGATTAG

>EucaDof4_Eucalyptus_grandis_CDS sequence_forward
ATGGATACTGCTCAATGGCCATCACAGGACACGAGGGTCGTCAAGGGAATGGAAGAGACA
TCTCCAAATGCATGCAATAATAGGGCCATGTTGGAGAAGAGAGCAAGGCCTCAGGAGCAG
CTGAATTGCCCAAGGTGTAATTCGAACAACACCAAGTTCTGCTATTACAACAACTACAGC
CTCACTCAACCCAGATACTTTTGCAAGACTTGCAGAAGGTACTGGACCGATGGCGGGTCT
CTAAGGAATGTCCCAGTGGGAGGCGGTTCGAGGAAGAACAAGAAAGTATCCACCTCGTCA
ACGTCGTCGTCACCGTCAGCTTCTTCTAGGCTCCCCGATCTTAACCCATCATCTGTAATC
TCTCACTTCTCTGCTCAAAACCCTAAGGCAGGACTACTGTACGAAGGTCAAGATCTCAAT
TTGGGCTTTCCAGGCACCCAAGAATACCGCGCAACAATTGCTCAGTATGTTCAAGTTCCC
AAAGTTGATCCTGAGACTAGTAGTAGAGGTCCGGATAACAACTCTAGTGCACCTTCTTCA
GCTCTAGAGCTGCTCCGGAATAGTTCTAGTGGAATGGGTTCAAGGGGCTTGAGTCTTCCT
TTCCTCCCAGCGAACATTCAGGCCGATTCCAATGTTCTCTACTCATCGGCAGGCTTTAAT
TTGCAAGAACTCAAGCCCTCGCTTGGCTTCTCTATGGATAGCAGGTATGCGAGTGCCCAT
GGAGTCGTTCATCAAGAGAATGGTGGGAAGTTACTGTTCCCCTTTGGAGGAATGATGAAG
CAGCTTTCGAGCACGAGTGGCAGCGATCAAGTAGATCACAATGGAAAGGGTCAGAGCAGT
TCGAATGCATATTGGAATGGAATGTTAGGTGGAGGTTCGTGGAAGGAAAAGTTGGACTCA
GAGGTCAATAGATACCTAAGGATTATTAGGAGTGTCTTCCTGCCTGCTGAGTCTTTTGGT
AAATACCCTGGACGATATCATGACGTTATTAAAGGACGTTTCTCAAGTATCAATGAGCTG
AGCATGGTATCGAAGACTTTAATAATATCTGTGAATTTGTGTTTCTACGGCTTTGTGGAT
TTCGATGTGCGGAAATTCTACTGTGTTGCTAAGCTGCTGTTGCATGTGCCTGTCATTGAG
CTCATCGGTGAACCTGATTAG

>EucaDof4_Eucalyptus_grandis_protein sequence
MDTAQWPSQDTRVVKGMEETSPNACNNRAMLEKRARPQEQLNCPRCNSNNTKFCYYNNYS
LTQPRYFCKTCRRYWTDGGSLRNVPVGGGSRKNKKVSTSSTSSSPSASSRLPDLNPSSVI
SHFSAQNPKAGLLYEGQDLNLGFPGTQEYRATIAQYVQVPKVDPETSSRGPDNNSSAPSS
ALELLRNSSSGMGSRGLSLPFLPANIQADSNVLYSSAGFNLQELKPSLGFSMDSRYASAH
GVVHQENGGKLLFPFGGMMKQLSSTSGSDQVDHNGKGQSSSNAYWNGMLGGGSKEKLDSE
VNRYLRIIRSVFLPAESFGKYPGRYHDVIKGRFSSINELSMVSKTLIISVNLCFYGFVDF
DVRKFYCVAKLLLHVPVIELIGEPD+

>EucaDof5_Eucalyptus_grandis_genomic sequence_reverse
ATGGACAACGCAGCCCCATCTCAGTGGACTGCACAGGTACTCTCTCTCTCTTCCCCCAAT
ACAATCATTTTCGGCTCAAGATTTTCTCTGCTCGAGTTGGTAGTTGGGGGCTTTGTGGAT
AAAAACTTTTTCCTCTCTTTGTTCGATTGCTTTGTTCCTTCCCCCTTCGACTTCAAATCA
TAAAGAAAAAAGAGAGGGGGAGGAAAGAAAAAAAGTTACCTTTTCTCCCCCACGAGATGA
TGTTCAGCAAGCAAAGCCCTCCACTAAATACGCTTGAATATATGTTGCTGTCGTTTTGGA
GATCTTTCTCGAGTCCTTGAAAGTACTGCCGACTTTCTTCTTCTTCTTCTTCTTCTTCTT
CTTTTCTCTCTTTTTCTTCTGTCTTCCTTCTGAAAAGTTATGATAATTTGTTTGTTTTGC
TCTCTTCTGGCATAATCACTCTTCCAAAAGCCTTCAAAGCACACCTTCTTTGCTTTGGTT
TTCTGTCTTTTTGTTTGTTTGTTTCTGCTCTTATTCCCCTTGTCTGTCTCTATTCATTCT
GTTTTCTTCATGCCTTCCTTCCTTCCTTCCTATGGTTGTTTGAATTTCAGGGATTCCTTG
GGGTGATTAATAATAACCCCCCAATGGAAATAGCATCATCACCGAGGCCCGTGTCCGTGA
GCTTGCAAGAGATCAAGCCTACATGTGGAACCACAACAACGAGCAGGCCGCATAAGGATC
AAGCCCTGAACTGCCCCAGGTGCAGTTCAAGCAACACCAAGTTCTGCTACTACAACAACT
ACAGCCTCTCTCAGCCCAGATATTTCTGCAAGACCTGTCGGAGGTACTGGACTGAGGGTG
GGTCCCTCAGGAACGTCCCTGTGGGTGGCGGCTCCAGGAAGAACAAGAGACCCTCTTCGT
CATTTTCTTCTTCGCCATCATCGAAGATCATCAAGCTTCCTGATCTGGTGGTGGGCCCGT
CAAGTCTTCCTCAAAACCCTAAGACCCATGAAGCCCAACGAGACCTAAACTTAACATACC
CACCTGATACTACTCCTACTACTGATCAACAAAATCCCAAGAGTAGCAGTGGAATTATTG
CTTCTTCTTCAATGGGATTGAGCTTAAGCTCCTTCATGCCGCACATGCCTGTAGCATCTG
ACTCAAGCTCCAACTTTTATGCGGCCGGTGGCTTGAATTTCCAGCATCTCCTTGGTCTCA
AACCGACGACATCCTCATTCGGTAGTTTTTCTGGACATGGGTTTGAGAGTGGTGGCTTTC
TTGGGAGCCTCCAAGGGTTGCAAGAGAGCGGCAATGGTGACGACAGTAATGATGGTGCGA
AGCTGTTTTTTCCCGTGGAGGATGGTTTGAAGCATCAAGTGACCAGTACTCCTCGTGTTG
ATGGTCAGTTTGGTCTTCAAAGTAGAGGACAAGGAGATCATTCAAATGGGTATTGGAATG
GAGTGCTGGGCACTGGAGGATCATGGTGA

>EucaDof5_Eucalyptus_grandis_CDS sequence_reverse
ATGGACAACGCAGCCCCATCTCAGTGGACTGCACAGGGATTCCTTGGGGTGATTAATAAT
AACCCCCCAATGGAAATAGCATCATCACCGAGGCCCGTGTCCGTGAGCTTGCAAGAGATC
AAGCCTACATGTGGAACCACAACAACGAGCAGGCCGCATAAGGATCAAGCCCTGAACTGC
CCCAGGTGCAGTTCAAGCAACACCAAGTTCTGCTACTACAACAACTACAGCCTCTCTCAG
CCCAGATATTTCTGCAAGACCTGTCGGAGGTACTGGACTGAGGGTGGGTCCCTCAGGAAC
GTCCCTGTGGGTGGCGGCTCCAGGAAGAACAAGAGACCCTCTTCGTCATTTTCTTCTTCG
CCATCATCGAAGATCATCAAGCTTCCTGATCTGGTGGTGGGCCCGTCAAGTCTTCCTCAA
AACCCTAAGACCCATGAAGCCCAACGAGACCTAAACTTAACATACCCACCTGATACTACT
CCTACTACTGATCAACAAAATCCCAAGAGTAGCAGTGGAATTATTGCTTCTTCTTCAATG
GGATTGAGCTTAAGCTCCTTCATGCCGCACATGCCTGTAGCATCTGACTCAAGCTCCAAC
TTTTATGCGGCCGGTGGCTTGAATTTCCAGCATCTCCTTGGTCTCAAACCGACGACATCC
TCATTCGGTAGTTTTTCTGGACATGGGTTTGAGAGTGGTGGCTTTCTTGGGAGCCTCCAA
GGGTTGCAAGAGAGCGGCAATGGTGACGACAGTAATGATGGTGCGAAGCTGTTTTTTCCC
GTGGAGGATGGTTTGAAGCATCAAGTGACCAGTACTCCTCGTGTTGATGGTCAGTTTGGT
CTTCAAAGTAGAGGACAAGGAGATCATTCAAATGGGTATTGGAATGGAGTGCTGGGCACT
GGAGGATCATGGTGA

>EucaDof5_Eucalyptus_grandis_protein sequence
MDNAAPSQWTAQGFLGVINNNPPMEIASSPRPVSVSLQEIKPTCGTTTTSRPHKDQALNC
PRCSSSNTKFCYYNNYSLSQPRYFCKTCRRYWTEGGSLRNVPVGGGSRKNKRPSSSFSSS
PSSKIIKLPDLVVGPSSLPQNPKTHEAQRDLNLTYPPDTTPTTDQQNPKSSSGIIASSSM
GLSLSSFMPHMPVASDSSSNFYAAGGLNFQHLLGLKPTTSSFGSFSGHGFESGGFLGSLQ
GLQESGNGDDSNDGAKLFFPVEDGLKHQVTSTPRVDGQFGLQSRGQGDHSNGYWNGVLGT
GGSW*

>EucaDof6_Eucalyptus_grandis_genomic sequence_reverse
ATGGTGGAGATGACCAGCGACCCTGCCATCAGGCTCTTCGGCAAGAAGATCCCCCTGTCG
TACACCGGGGAGTCGCCGGAGGCCTCCTCCGGCGACGGCGTGGTTTACTCGGGCGTGAAT
TCGGAGGAGGACGACGAGGGCTCCGAGAAGGTACGGTTTGTCGGGTGTGGAGGTGTCTGT
GCAATTCGAGGACTGCCCATCTACTCTCTGTCTCTCTGTCTCTCTGTCTCTGCTTTTCTG
TACTTGGGGTGTGGTGCTCTGTCTCTGGATGATTCTGTGTAGCTTTTTCGAGATCTTGGA
ATTCTTTTTTGCTATTTTCAGCTTCTTTTTACGATGGTAGTCTGTGAATTGGTTTTTGAT
TTGATCCATGAAGTTGAATTCATAGAAAAGTTTAAAGTGAAGACGAGAGGAACTGACCAG
TCTTCGCGAGGTTAATTCTTTTGCTTGTTTTAAGGATCTTGTCTTCACTTAAAACAGCAT
AAAACGAAGACGGCTAACGACCCTGCTAATTTCATCCAGAAAAAGGGGGAAAAAACATAA
ATTTGACCAAAAAAATCAAAATTTTCAAGGAAGTATGGAAAGAAGGAAGAGATCTGTTTC
TTTTCCCTTTTTTCGTTTACTTTGTCTTTACCGAAATATGGGATGAATAATCTGTGTCTG
TATTGGTCAAGAAAGACCCTTTTGGTTTTCGCCCTAGTTAAAATTGATTGAAATCTTCAC
CTCTTCCCTGTTTAAAATACGGTATGGACAGGCCTTTTAGCTGTATACTTTCTGGTATGA
TCCATGATGAGCATAATTTATACTATGCGGATAGAGAAATTCCGCATACGCTCTTAAATT
CTAGTCCTAGACATCACATGTGAGCTGATAATGGCTTTTAGGACCAGCTGCATCGGTCCA
AGGCCGCATGTTTCGGATCTGTCACTATTTTAAGTCACTAAATCTTTGTCTTTTTCCTTA
ATCATGTCAAATTTCTTGGATAACTCATCCATGTCTCGCCGGTCCAGGCGACCGGATCGA
TGTCAAGTGTCACGTTACAGAACTTTCCAAGACTCTAGCTTTAGCTAAGGGCAATTAAAC
AATGGGAAGAGGGATGGCATTATTTTAGTCAAGAAATGATGGCAAAAAGCATTCTCTAGA
TGCTCTTCTCATTCTTTCCTGTTGCGTCTTGCTTGTTCTGTTGCTGGGCATCATTTTCTT
CTAGAATTTCCTAGGTGAGACTTTTACATCCGTTGCTCAGTGCTGATTTCTCTCCACTTT
GTCATGTCATCAGTGTCATTCCATTGAACACTTTGCCATGGAAGTGCAAGAAGATGATGA
CCAGGTTTCACCAGAAACGGTGAACCCCAAAACACCCTCTATTGAGGAAGAATCTGCTAA
AACGAAAGCTTCCGGCATCGATCAAGAACAGGGCGATTCGTCGAACTCGCAGGAGAAGCC
CCCCCTGAAGAAGCCGGACAAGATCATACCTTGCCCGCGATGCAACAGCATGGACACCAA
GTTCTGCTACTACAACAACTACAACGTCAATCAGCCCCGGCACTTCTGCAAGGCCTGCCA
AAGATACTGGACGGCCGGCGGTACCATGCGGAACGTTCCCGTGGGAGCTGGGCGCCGCAA
GAGCAAGAGCTCAGCTTCGCATTATCGCCAGATCACTCTCTCTGAGGCTCTTCAGGCAGC
TCGAATTGATGCTCCGAACGGGACTCACCTCCCAAATGGGAAGATCAATGGTACTGTTCT
CACCTTCGGCTTAGATCCATCTGCTGGTGATTCCATGGCTGTTGCCTTGAATCTCTCCAA
TAAGGAAGGATCAGCTATGAACAGAATGCAGAATGGGTTTTGTGACTTCGGAAGGAATGG
GGTTTCGGTCCCTCAAAGAAGTGGAGAGAAGAATGACGATTGCTCTAGCCGATCTTCTGT
CGCGATCTCGAACTTGACAGATGAAGGAAATGGGAATTCCAATCTGGAGGGATCCAGTGC

CAATGGATTCCATCCTCAGTTTCCATGCTATCCGGGTGTTCCACAGCCTTATCCATGGAA
TCCTGCGATGCTTCCACCACCCTTTTATCCTTCAGGATTTGCTGTGCCGATTTACCCTGC
GCCGCTTTGGAACTGTGGCCTCCACAGTAGCTGGAATTTCCCTCTGTTGTCGCCTCCGAG
CCAAAAGGGATCAAATTCAGGACCAGATTCGCCGACTCTTGGGAAGCATGCAAGAGAAGG
CGAGATGGTTGAAAAGGTTGATTCTGTGGAAGACGAGCCCAGGAAACGGAAGAACGGATG
TGTTTTGGTCCCGAAGACTTTACGCATCGATGACCCAAGCGAGGCAGCGAAAAGTACTAT
ATGGGCTACTCTGGGGATCAAGAAGGACTGGCTAGGAAAGGACGGGTTGTTTCAGGGCTT
CCAATCGAAGGGCGAGGAGAAAAATCACGTATCAGAAACCGCATCTGTTTTGTGCGCTAA
CCCTGCAGCCTTATCGAGATCCATCAAATTCCACGAGAGCTCTTGCGTCTAGCTCATGAT
AACTACAGGTAGAAATAACTACAGAAAACTCTTGAAACTGGGCTGTCATCTTTCTTTGGTG
ATTGCAGTGACTTGCATTGTTTAGAACAGAGAAAGGCAAGACATTTGATGGCCGAGACAAG
CAGAGGATAGCGATAGACCGTGACTTGCTGGAGTTTGCAGCCACAGTTAGTTTTAACCCGA
TGGGGTTGTTTCCTTTTCTTGTGTGCTAATTAGGTTGAATGCACGCGCTTGTACATAACTC
TTGTAAATTATTCGGACGAGATGGAGAGAACTTTTTTCTGAATTTTGTACCAGTTCTTCAT
GTCTCTTGAGAGATGTTTGTTAAGTGCATCCTGAATCAAAACAGTGATGTTTTAAACCTCT
TCGATTTCTCCTAGTTTTCTTTGTAATATTCCAACAAAGAATATCATCAAGATGTTCTTGA
GATGATGTTCAGACTAAATATGATTGAATCGCTCAAGGATGACAGAACTGAAGCAGGGCAC
AGCTTTGGGGGAAGAAATATACCATGAAGACTTGGGCAGAAAAAGCCTACTTTTTCTAGAT
CCCAAGACCGATGGGTATCTAGTAAAAGCAACAATGTTTCTTTCCTTGCAATCCAAAACAC
TAGTTGTCACAGTTTGGCAATGAAAGGTCAGTACTTTCACTGTTTCTTTGGTCTATTTTTT
CAAAAGTGATATCGAGAAACAGCTAGTAGCAAGGCATGGTTTTAACAATGCTCACTGA

>EucaDof6_Eucalyptus_grandis_CDS sequence_reverse
ATGGTGGAGATGACCAGCGACCCTGCCATCAGGCTCTTCGGCAAGAAGATCCCCCTGTCG
TACACCGGGGAGTCGCCGGAGGCCTCCTCCGGCGACGGCGTGGTTTACTCGGGCGTGAAT
TCGGAGGAGGACGACGAGGGCTCCGAGAAGTGTCATTCCATTGAACACTTTGCCATGGAA
GTGCAAGAAGATGATGACCAGGTTTCACCAGAAACGGTGAACCCCAAAACACCCTCTATT
GAGGAAGAATCTGCTAAAACGAAAGCTTCCGGCATCGATCAAGAACAGGGCGATTCGTCG
AACTCGCAGGAGAAGCCCCCCCTGAAGAAGCCGGACAAGATCATACCTTGCCCGCGATGC
AACAGCATGGACACCAAGTTCTGCTACTACAACAACTACAACGTCAATCAGCCCCGGCAC
TTCTGCAAGGCCTGCCAAAGATACTGGACGGCCGGCGGTACCATGCGGAACGTTCCCGTG
GGAGCTGGGCGCCGCAAGAGCAAGAGCTCAGCTTCGCATTATCGCCAGATCACTCTCTCT
GAGGCTCTTCAGGCAGCTCGAATTGATGCTCCGAACGGGACTCACCTCCCAAATGGGAAG
ATCAATGGTACTGTTCTCACCTTCGGCTTAGATCCATCTGCTGGTGATTCCATGGCTGTT
GCCTTGAATCTCTCCAATAAGGAAGGATCAGCTATGAACAGAATGCAGAATGGGTTTTGT
GACTTCGGAAGGAATGGGGTTTCGGTCCCTCAAAGAAGTGGAGAGAAGAATGACGATTGC
TCTAGCCGATCTTCTGTCGCGATCTCGAACTTGACAGATGAAGGAAATGGGAATTCCAAT
CTGGAGGGATCCAGTGCCAATGGATTCCATCCTCAGTTTCCATGCTATCCGGGTGTTCCA
CAGCCTTATCCATGGAATCCTGCGATGCTTCCACCACCCTTTTATCCTTCAGGATTTGCT
GTGCCGATTTACCCTGCGCCGCTTTGGAACTGTGGCCTCCACAGTAGCTGGAATTTCCCT
CTGTTGTCGCCTCCGAGCCAAAAGGGATCAAATTCAGGACCAGATTCGCCGACTCTTGGG
AAGCATGCAAGAGAAGGCGAGATGGTTGAAAAGGTTGATTCTGTGGAAGACGAGCCCAGG
AAACGGAAGAACGGATGTGTTTTGGTCCCGAAGACTTTACGCATCGATGACCCAAGCGAG
GCAGCGAAAAGTACTATATGGGCTACTCTGGGGATCAAGAAGGACTGGCTAGGAAAGGAC
GGCCTTATCGAGATCCATCAAATTCCACGAGAGCTCTTGCGTCTAGCTCATGATAACTAC
AGAACAGAGAAAGGCAAGACATTTGATGGCCGAGACAAGCAGAGGATAGCGATAGACCGT
GACTTGCTGGAGTTTGCAGCCAAACTGAAGCAGGGCACAGCTTTGGGGGAAGAAATATAC
CATGAAGACTTGGGCAGAAAAAGCCTACTTTTTCTAGATCCCAAGACCGATGGTGATATC
GAGAAACAGCTAGTAGCAAGGCATGGTTTTAACAATGCTCACTGA

>EucaDof6_Eucalyptus_grandis_protein sequence
MVEMTSDPAIRLFGKKIPLSYTGESPEASSGDGVVYSGVNSEEDDEGSEKCHSIEHFAME
VQEDDDQVSPETVNPKTPSIEEESAKTKASGIDQEQGDSSNSQEKPPLKKPDKIIPCPRC
NSMDTKFCYYNNYNVNQPRHFCKACQRYWTAGGTMRNVPVGAGRRKSKSSASHYRQITLS
EALQAARIDAPNGTHLPNGKINGTVLTFGLDPSAGDSMAVALNLSNKEGSAMNRMQNGFC
DFGRNGVSVPQRSGEKNDDCSSRSSVAISNLTDEGNGNSNLEGSSANGFHPQFPCYPGVP
QPYPWNPAMLPPPFYPSGFAVPIYPAPLWNCGLHSSWNFPLLSPPSQKGSNSGPDSPTLG
KHAREGEMVEKVDSVEDEPRKRKNGCVLVPKTLRIDDPSEAAKSTIWATLGIKKDWLGKD
LIEIHQIPRELLRLAHDNYTEKGKTFDGRDKQRIAIDRDLLEFAALKQGTALGEEIYHED
LGRKSLLFLDPKTDGDIEKQLVARHGFNNAH*

>EucaDof7_Eucalyptus_grandis_genomic sequence_forward
ATGGAGAGAGCATGGAACCCAAGTGTTGAGATGTCACCCAACTGTCCACGTTGTGGATCT
TCCAACACCAAGTTCTGCTACTACAACAACTACAGCCTCACTCAGCCCAGGTACTTCTGC
AAAGGTTGCCGGAGGTACTGGACGAAAGGTGGCTCTCTCCGGAATGTCCCCGTCGGCGGC
GGTTGCCGGAAAAACAGGCGGGGCAAGACCATACGTATATCCTCCAGCGCTGCTACTTCA
TCAGCTTACCTTGCCGCTGCCAGCAGCCCTTTTGGGAGTGGATTGTCGTGTGGTTCTATG
TTGAGAGACAGCAATAGTTCCTTGGCCGATCCAAGTAGCTCGTCTAATGGCAGCTCGCAT
ATTGATCTTGCCCTCGTCTATGCGAATTACTTGAATCAGAAGCCCGAACCGGCACACGAA
GAGTTGGCCGGTGGTGAATCCGATCACCACGACAGTCAGCCTCTGGCAAATTCAAAGCAA
GTTCACGACTTGGACAATGTCGAATTGATGCATGGAAATGAGCTTCTTGGATGCGAGTTC
ATTTCTGAGCTTTTTTCTGAGACTCACTTGAGCAATGGAAATAACCAGATGTACCATCAA
CCTCATCATCAAGCAGGTGACTCTAACAATCTTAACGGGATCATTGATCAGATTCAGTCG
TGTGATGGTGGTGATGCGAGCTACTTTGGCGGGTTGCCACCTTTGCCAAGCCAAGGGGGC
GGCACGTGGTTGAATGACCCTGACCATTTGATGATGAATCATGGGTTTGCGAGCACAGAG
CTCATGCCAGTTCTTGGTCCAGAGGCTAATCCAGATCTCTTGGCCAGTAACTGGAGGGAA
TTTGATACGTCTATACAGGATGCTACTTTCTCCGGGAATTGA

>EucaDof7_Eucalyptus_grandis_CDS sequence_forward
ATGGAGAGAGCATGGAACCCAAGTGTTGAGATGTCACCCAACTGTCCACGTTGTGGATCT
TCCAACACCAAGTTCTGCTACTACAACAACTACAGCCTCACTCAGCCCAGGTACTTCTGC
AAAGGTTGCCGGAGGTACTGGACGAAAGGTGGCTCTCTCCGGAATGTCCCCGTCGGCGGC
GGTTGCCGGAAAAACAGGCGGGGCAAGACCATACGTATATCCTCCAGCGCTGCTACTTCA
TCAGCTTACCTTGCCGCTGCCAGCAGCCCTTTTGGGAGTGGATTGTCGTGTGGTTCTATG
TTGAGAGACAGCAATAGTTCCTTGGCCGATCCAAGTAGCTCGTCTAATGGCAGCTCGCAT
ATTGATCTTGCCCTCGTCTATGCGAATTACTTGAATCAGAAGCCCGAACCGGCACACGAA
GAGTTGGCCGGTGGTGAATCCGATCACCACGACAGTCAGCCTCTGGCAAATTCAAAGCAA
GTTCACGACTTGGACAATGTCGAATTGATGCATGGAAATGAGCTTCTTGGATGCGAGTTC
ATTTCTGAGCTTTTTTCTGAGACTCACTTGAGCAATGGAAATAACCAGATGTACCATCAA
CCTCATCATCAAGCAGGTGACTCTAACAATCTTAACGGGATCATTGATCAGATTCAGTCG
TGTGATGGTGGTGATGCGAGCTACTTTGGCGGGTTGCCACCTTTGCCAAGCCAAGGGGGC
GGCACGTGGTTGAATGACCCTGACCATTTGATGATGAATCATGGGTTTGCGAGCACAGAG
CTCATGCCAGTTCTTGGTCCAGAGGCTAATCCAGATCTCTTGGCCAGTAACTGGAGGGAA
TTTGATACGTCTATACAGGATGCTACTTTCTCCGGGAATTGA

>EucaDof7_Eucalyptus_grandis_protein sequence
MERAWNPSVEMSPNCPRCGSSNTKFCYYNNYSLTQPRYFCKGCRRYWTKGGSLRNVPVGG
GCRKNRRGKTIRISSSAATSSAYLAAASSPFGSGLSCGSMLRDSNSSLADPSSSSNGSSH
IDLALVYANYLNQKPEPAHEELAGGESDHHDSQPLANSKQVHDLDNVELMHGNELLGCEF
ISELFSETHLSNGNNQMYHQPHHQAGDSNNLNGIIDQIQSCDGGDASYFGGLPPLPSQGG
GTWLNDPDHLMMNHGFASTELMPVLGPEANPDLLASNWREFDTSIQDATFSGN*

>EucaDof8_Eucalyptus_grandis_genomic sequence_forward
ATGCCGTCGGAGTGCGCCGACCGCAAGCCCGCGAAGGCCCCCCACGGCGGCCACCTCCAC
GCGGCCGGGGCGGGGGCCGCCGCGCCGCAGCCGCAGGAGCAGGAGCACCTCCCCTGCCCG
CGCTGCGACTCCACCAACACCAAGTTCTGCTACTACAACAACTACAACTTCTCCCAGCCC
CGCCACTTCTGCAAGTCCTGCCGCCGCTACTGGACCCACGGCGGCACCCTCAGGGACATC
CCCGTCGGCGGCGGCAGCCGCAAGAACGCCAAGCGCTCCCGCACCGGCGCCTCCGCGTCG
TCGTCGTCCTCCTCCGCCGCCGTCGTGCTCCCCGGCGTCGCCCCCCAGGAGCACCACTTG
CACGCCGCCTCCCTCCTCCTGCCCTTCGCCGGGAGCCACGGGGGGCCCGTACACCTCGCC
GGCGGCGACGGCGACGGGAAGCCGGGCCTGAACGTGTGCGGGAGCTTCACTTCCCTGTTG
AACGCTCACGGGCCTCCTGGGTTCCTGGCTCTGGGCGGGTTCGGGCTCGGGCTCGGGCCG
GGCGCGCTCGACGAGATGGCGTTCGGGCTCGGGAGAGGGGTCTGGCCCTTTCAGGGGGTG
GGAGACGGCATCGGCGCGGGCGTGGGGAGCAATGGCGGGCCAGCGGGGATCGGGAACGCG
TGGCAGTTCGAGGGCGGCGACGAGGGCGGGCTCGTGGGCGGCGGCGGCGGAGGCGGCGAC
TCGTGCTTCTCTTGGCCAGGACTCGCCATCTCCACCGCTTGGAGCGGTCTCAAATGA

>EucaDof8_Eucalyptus_grandis_CDS sequence_forward
ATGCCGTCGGAGTGCGCCGACCGCAAGCCCGCGAAGGCCCCCCACGGCGGCCACCTCCAC
GCGGCCGGGGCGGGGGCCGCCGCGCCGCAGCCGCAGGAGCAGGAGCACCTCCCCTGCCCG
CGCTGCGACTCCACCAACACCAAGTTCTGCTACTACAACAACTACAACTTCTCCCAGCCC
CGCCACTTCTGCAAGTCCTGCCGCCGCTACTGGACCCACGGCGGCACCCTCAGGGACATC

CCCGTCGGCGGCGGCAGCCGCAAGAACGCCAAGCGCTCCCGCACCGGCGCCTCCGCGTCG
TCGTCGTCCTCCTCCGCCGCCGTCGTGCTCCCCGGCGTCGCCCCCCAGGAGCACCACTTG
CACGCCGCCTCCCTCCTCCTGCCCTTCGCCGGGAGCCACGGGGGGCCCGTACACCTCGCC
GGCGGCGACGGCGACGGGAAGCCGGGCCTGAACGTGTGCGGGAGCTTCACTTCCCTGTTG
AACGCTCACGGGCCTCCTGGGTTCCTGGCTCTGGGCGGGTTCGGGCTCGGGCTCGGGCCG
GGCGCGCTCGACGAGATGGCGTTCGGGCTCGGGAGAGGGGTCTGGCCCTTTCAGGGGGTG
GGAGACGGCATCGGCGCGGGCGTGGGGAGCAATGGCGGGCCAGCGGGGATCGGGAACGCG
TGGCAGTTCGAGGGCGGCGACGAGGGCGGGCTCGTGGGCGGCGGCGGCGGAGGCGGCGAC
TCGTGCTTCTCTTGGCCAGGACTCGCCATCTCCACCGCTTGGAGCGGTCTCAAATGA

>EucaDof8_Eucalyptus_grandis_protein sequence
MPSECADRKPAKAPHGGHLHAAGAGAAAPQPQEQEHLPCPRCDSTNTKFCYYNNYNFSQP
RHFCKSCRRYWTHGGTLRDIPVGGGSRKNAKRSRTGASASSSSSSAAVVLPGVAPQEHHL
HAASLLLPFAGSHGGPVHLAGGDGDGKPGLNVCGSFTSLLNAHGPPGFLALGGFGLGLGP
GALDEMAFGLGRGVWPFQGVGDGIGAGVGSNGGPAGIGNAWQFEGGDEGGLVGGGGGGGD
SCFSWPGLAISTAWSGLK*

>EucaDof9_Eucalyptus_grandis_genomic sequence_reverse
ATGCCTTCGGACGCCCGGCCGCAGGCCTCCGCCGCTGCAGCCGCCGGGGGAGCTGGCCTG
CACCCCCACCCGCCGCCGCAGACGGAGGCCCTGCCCTGCCCGCGCTGCGACTCCACCAAC
ACCAAGTTCTGCTACTTCAACAACTACAACCTCTCCCAGCCCCGCCACTTCTGCAAGTCC
TGCCGCCGCTACTGGACCCAGGGCGGCACCCTCCGCAACGTCCCCGTCGGCGGCGGCTCC
CGCAAGGCGTCCTCCTCCTCCTCCTCGAAGCGGCCCCGCTCGCTGGCCCCGCCTACTTCT
TCGTCTCCTTGCTCAAACTCCAATTCCTCCTCCGTGTCCCAGGAGCCCCTGCACGGGGAC
CACTCGGGCCTGCTCCTGGCCCAGGTGAAGTCCGAGGTGGGCTCGGGCGATCAGGTGAAG
AGGACCGAGAATGTGGCGGGACCCGGTGTCGGCTCCGGCCACCAGGGTTTCACTTCTCTG
CTCGGCTCGCCGGGGGGCCAGGGGTTTCTGGCTCTGGGCGGGCTGGGGCTCGGGACGGGG
CTGGACGGCCTAGGGTTCGGGTTCGGGAGGGGGGTCTGGCCGTTCAACGAGGTCGGCTTC
GTCGACGGCGGCGGCGTGTATGGTGGCGGCGGCGGCGGGGGGCCCGTGGCGGCGGCGGGA
GGCAACGCGTGGCAGATGGGGAGCGGGGAAGGCGGAGGGCTGGTGGACGGCGACGGGTTC
GCTTGGCCGGAGCTCGCGATATCGATGCCGGGTCAGGGCCTGAAGTGAGTAGCGTGGCTC
GTCATTTTGCTGTGCTTTTCCCGTGTCGGGTTTGGGAGTTGGTGGGCGAGGGTCAAGAGT
CGAGGGTGGCATCAGGGTTTCCTTTTTCTTTTCTCTTTTCCGGGTGGGGGATTTAGGTTG
GGGATATAAGTATTCACTGAATAGGTTTTAAATTACAGGTTACATAGGAAGATATCGCGA
TGACCACCGGAATGATATGATGGTTTAGCGTAGGTGAACGATGTTCGGTCGGTCTTTTGT
TTTCCACTCTGCGGTTTGGTCCATGACTGCCCTGTGTAAATCTCGAATTTGGACATCGAA
ATTGAAGTTAATGACTTGTTTTGGCTTGATGCGAGCTCTCGCTACTTGTGTTTAGTTTCT
CTTTTGTTGGGGGTTTAAATTTGGTTTGGTGATGGCTGGAATAGACAGATGGAAGGGCAG
AGAGTGTAACATTTTACACGGACGGACGAGTTAGCGTTTGATTCATTCGATGGCACGGAG
AAATCGCTCTCGCATCACGGCTTAGGCCTGCCAATCGCGAGGTTAATAGAGGGTATGTTG
ACGCTCGAGGAGAAGCGTGTCGATCGATTTCAACTCGCCGACCCTTCGATTTGGTTTAGC
GTCTGTCGAGTGATGCATATGCTAGCGGTGTGAGTTGAGTACCGAAAGTAGTTCATAGCT
TGATGATGCTTGATATGATATATATAAGGCGGGACTGGTCTGGGCGAAGCTCTGCGACTT
GCACTCTTCAAACGCTTAGAAAAGGAAGCGTCGCGATAACGAATGCATTTGCGCGGCTAA
TGCAGTTTCGGCAAAATCACCTTGCCAACCTTGCTAACGGTCCACTTACGTCTGTCTCGT
CGCTAGAGCGAATCTTGCTAACCTTGCTAACGGTCCACTTACGTACATAATCTTCCATCG
CGTTGCTAGAGCGAGAATCTCGTTAATCATCTTTTTGGACGAAACCATGTTGCCAAATAA
ATGTGGAAAGTGCGCATCTATGCATATGTGCCGACGGTATTCTCCGTCGTGGGATATCAT
TCATTTCCTCGTCATCATCCAATTGGTATTTATCGATACACGTTCAATGCATTCTCGTCC
CGGGTGGATCCGCCATTGATTGTTTTGGGCCTACCCACCCTCGAAGCATTGATTCTTCCA
CGACCTTATTATGTTGTATCTCACGGGCGCCAATGCACGGTTCGAATCGGAATAGAGAAT
TTCGAGTGATGATTGCTTTGACACCTGAAGAAGCATCCATGATGATTGCGCATCATCTCC
CCCACCAAACGCACATGATTCGATTGTGTAAACGGTCATCATAGGATAAGCATTGAATGT
ATAACGAGTATATAGCCTCGACATGGTGCGATTTAACCCAAAGTGACAACTTCCACGACC
ACTCTTGCATAGCATCATCTCCTATTGAGTCGCCATCTTCGGGTGTAGTGAGATACAGGG
ATCACACTCCGTGGTGAGCGGATGGGGGTCTTGGGACTACCTCCGCGAAGTTGCAAGAGT
AGTGTCCTTGAGTCGTCAGTAGGCTTTTATAATTTGAGTCAATATTTTATTAGTAGTTTG
GATTTGGGTCTATAAATAGTTGCTGGTATATATAATCTCTCTCTTTTTTTTTTCTTTTTC
TTTTTTAATAATTGAGAGATGTAATAAAGAGGTGTATTTTGCTAATTATAGTGAAATCCT
TTACTGAATATAACCTTAATTTAAGAGTGAATCAAGATAAAACCTTATCTCTTGAATTTT
CTCTCTCACTCTACTTTCTAATTCCTTGTTATTGTGCATTGAATCCCAACACTTTAGATG

TTTAAAAGAAGACCGGTGCTCTCACTCATTAACCTTCTAATATCTCAGTTTTGCTCGCAA
GTTGTCCGGAAATCAAACCTCCGCGTGGCAATTATTGCGATTAAGAGGGTGCCGACTCAG
GACCGTGTAAGTAAGTGTGAACATGATTTTAGGATAGAATCGGCAACTGAAGAACGAGAC
TGGTGGGAATAGCGCTCAGCTTCAAGATTTTGGGAGCCTTGGTCCAGCTCTCGATTCTAT
CTTTCGGGGAACCGAGAGCCGATAATGCAGACTAGACGCATACGACAGACTGTACCTGCA
AGAGCAGAGGGACCAAGAGGCGCAGCGATAAATGTCTAAGAAGGCCATCACCACGGCAAA
ATTTCGCTGCTGTCACGATCAACATGCAGGTGTAATTTCCTTCGAACAGCGACAGCGGAA
GTTTCTGCGTAAGAACCCTGTTTACTTTGCTACAGTTAAAAGCTCATATCGAGGAGTCAA
TCGCCACCTTCTTCTCAAATCGAGGAGTCAATGACCACCTTCTTCCCAACTGTTCAGAGA
TATCGCTGGCGAGAGGGGACCTTCTCTTTGCTTGCGAAAGCAAGTGATGCCAGCCAGCCT
TCCTCGACAAGACAGATCCCACCGTGCAATGCAACCCATCTCAGCAGGATTCATTTGGCT
ACAAGACACGAGTATCCAAAGAGAACATTTTAGGAGTGAGATTGCGATCATCGGAAATGA
GATCTGCGAGTAAGATTGCTAAGGCTGCGTTTGTTTGCAACTTGTTTCTGTTTCTAGGAA
CAAAAATTTCTGTTTTTTTTTGTTCCTGTGAACAAAAAAGAAACAAAAACGCGTTTGGGTG
CGTTTTTGTTCCAAAATTATTTTTTTGTTCCCAGAAACACATCTGGAACAGAAACAAGAA
ACAAAAAAAAGTTATTTATTGTTTCTGGAAAACATTTTAGAAACACTTTCTTTTCTTATT
CTTTTTTCTTTCTTTATTCTTTGGCCGGGGAGGGCCGGCGACCTTGCCGGAGCGTCGCCG
ACCCTCGACAAGGCCGAGCTCGCCCAGTGGCCGGGGGGAGCGGCTGCAAGTGAACGGCGG
CGGGCTCCAACTCAGGGCAGTGGCGGAGTCGGGGTTGGGTCGCGGGTCGTTGGGCGGCGG
CGGCGGCTGCATTGCGGGGCTCGGCCGTGGGATCTCGAGTGGCGGACGTCAAGGAGCGAA
GCGACGAGGACGGCCGGTCGAGTCGTGGGCCAGCGCGGGGTCGATGATCGAGCGGATCTC
GAAGTCGTGGGTCACGGCGCGGTGAGATCTGATGCGGCGGTGGCGAACGGAGGCTCGAGA
CCCGCAGATGCGTCGATCGGGCGTGAGGCGGAGGGATGGTCTTTGGGGCTTCGGCGTCGC
TAGGTCGTCGGCGGAGGCTTCGGGTGCAGGTCTGGTGTCACGAGGGGAGAACACGGAGCA
GCAGGCGACGGATCTGGGGACGCAGAGGTTGGCGACGCCGTCGGGCGCGAGTCCGGCGTC
GCGGCGTGCAGGGGGCGCGGTGGCTGCTCCGACGAACGGCTGGTACGGTGGCGGCGAGGA
GGCGTCGAGCGCAATCGCCGATGCGGTGGAGGGAGACGACAGTGTGGCTCTGGGTCACCG
TCTGCTCGACTGTCGGCGCGGGCGCGCTGGCACAGTCGTGCGAGGAGACGCCGGGGAGGA
AGACCGACTAGCCAAAAGAAGAAAAAGGAAAAAAGGAAAAAAGTAA

>EucaDof9_Eucalyptus_grandis_CDS sequence_reverse
ATGCCTTCGGACGCCCGGCCGCAGGCCTCCGCCGCTGCAGCCGCCGGGGGAGCTGGCCTG
CACCCCCACCCGCCGCCGCAGACGGAGGCCCTGCCCTGCCCGCGCTGCGACTCCACCAAC
ACCAAGTTCTGCTACTTCAACAACTACAACCTCTCCCAGCCCCGCCACTTCTGCAAGTCC
TGCCGCCGCTACTGGACCCAGGGCGGCACCCTCCGCAACGTCCCCGTCGGCGGCGGCTCC
CGCAAGGCGTCCTCCTCCTCCTCCTCGAAGCGGCCCCGCTCGCTGGCCCCGCCTACTTCT
TCGTCTCCTTGCTCAAACTCCAATTCCTCCTCCGTGTCCCAGGAGCCCCTGCACGGGGAC
CACTCGGGCCTGCTCCTGGCCCAGGTGAAGTCCGAGGTGGGCTCGGGCGATCAGGTGAAG
AGGACCGAGAATGTGGCGGGACCCGGTGTCGGCTCCGGCCACCAGGGTTTCACTTCTCTG
CTCGGCTCGCCGGGGGGCCAGGGGTTTCTGGCTCTGGGCGGGCTGGGGCTCGGGACGGGG
CTGGACGGCCTAGGGTTCGGGTTCGGGAGGGGGGTCTGGCCGTTCAACGAGGTCGGCTTC
GTCGACGGCGGCGGCGTGTATGGTGGCGGCGGCGGCGGGGGGCCCGTGGCGGCGGCGGGA
GGCAACGCGTGGCAGATGGGGAGCGGGGAAGGCGGAGGGCTGGTGGACGGCGACGGGTTC
GCTTGGCCGGAGCTCGCGATATCGATGCCGGGTCAGGGCCTGAAGTGTAATTTCCTTCGA
ACAGCGACAGCGGAAGTTTCTGCAGATATCGCTGGCGAGAGGGGACCTTCTCTTTGCTTG
CGAAAGCAAGTGATGCCAGCCAGCCTTCCTCGACAAGACAGATCCCACCGTGCAATGCAA
CCCATCTCAGCAGGATTCATTTGGCTACAAGACACGAGTATCCAAAGAGAACATTTTAGG
AGTGAGATTGCGATCATCGGAAATGAGATCTGCGAGGCCGGCGACCTTGCCGGAGCGTCG
CCGACCCTCGACAAGGCCGAGCTCGCCCAGTGGCCGGGGGGAGCGGCTGCAAGTGAACGG
CGGCGGGCTCCAACTCAGGGCAGTGGCGGAGTCGGGGTTGGGTCGCGGGTCGTTGGGCGG
CGGCGGCGGCTGCATTGCGGGGCTCGGCCGTGGGATCTCGAGTGGCGGACGTCAAGGAGC
GAAGCGACGAGGACGGCCGGTCGAGTCGTGGGCCAGCGCGGGGTCGATGATCGAGCGGAT
CTCGAAGTCGTGGGTCACGGCGCGGTGAGATCTGATGCGGCGGTGGCGAACGGAGGCTCG
AGACCCGCAGATGCGTCGATCGGGCGTGAGGCGGAGGGATGGTCGTCGGCGGAGGCTTCG
GGTGCAGGTCTGGTGTCACGAGGGGAGAACACGGAGCAGCAGGCGACGGATCTGGGGACG
CAGAGGTTGGCGACGCCGTCGGGCGCGAGTCCGGCGTCGCGGCGTGCAGGGGGCGCGGTG
GCTGCTCCGACGAACGGCTGGTACGGTGGCGGCGAGGAGGCGTCGAGCGCAATCGCCGAT
GCGGTGGAGGGAGACGACAGTGTGGCTCTGGGTCACCGTCTGCTCGACTGTCGGCGCGGG
CGCGCTGGCACAGTCGTGCGAGGAGACGCCGGGGAGGAAGACCGACTAGCCAAAAGAAGA
AAAAGGAAAAAAGGAAAAAAGTAA

>EucaDof9_Eucalyptus_grandis_protein sequence
MPSDARPQASAAAAAGGAGLHPHPPPQTEALPCPRCDSTNTKFCYFNNYNLSQPRHFCKS
CRRYWTQGGTLRNVPVGGGSRKASSSSSSKRPRSLAPPTSSSPCSNSNSSSVSQEPLHGD
HSGLLLAQVKSEVGSGDQVKRTENVAGPGVGSGHQGFTSLLGSPGGQGFLALGGLGLGTG
LDGLGFGFGRGVWPFNEVGFVDGGGVYGGGGGGGPVAAAGGNAWQMGSGEGGGLVDGDGF
AWPELAISMPGQGLVISFEQRQRKFLDIAGERGPSLCLRKQVMPASLPRQDRSHRAMQPI
SAGFIWLQDTSIQREHFRSEIAIIGNEICEAGDLAGASPTLDKAELAQWPGGAAASERRR
APTQGSGGVGVGSRVVGRRRRLHCGARPWDLEWRTSRSEATRTAGRVVGQRGVDDRADLE
VVGHGAVRSDAAVANGGSRPADASIGREAEGWVVGGGFGCRSGVTRGEHGAAGDGSGDAE
VGDAVGRESGVAACRGRGGCSDERLVRWRRGGVERNRRCGGGRRQCGSGSPSARLSARAR
WHSRARRRRGGRPTSQKKKKEKRKKV#

>EucaDof10_Eucalyptus_grandis_genomic sequence_forward
ATGGTTTTCTCATCCCTTCCAGCTTATCTTGATCCATCCAACTGGCAACAGGTGAGGGTT
TTCTTCAAACATGGTGTATTTGCTTTGCAAGAAAAAGCTAAATTTTTTCAAAAGATAAAT
ATTCTCATGTGTTACTTTTGTTAATGCCTTGTTTGAATATACCACAGATATTTGATTTCT
TGCTTTTATTCCACACTTTTCACAGCAACAAAATCATCACCATCATCAAGCTGGAAATAG
TGGTCTCAACAATCAGCTCCCACCTGCTCCTCCGCCGCCACCGCCTCCTCCTCCACAACC
TCATGGAGGAGCTGGAGGCAACCCTGGCGGCTCGATCCGTCCGGGCTCGATGGTTGAGCG
GGCCCGGCTAGCCAACATACCCATGCCAGAGGCAGCGCTCAAGTGCCCGCGATGCGATTC
CGCCAACACGAAGTTTTGCTACTTCAACAACTACAGCCTCTCTCAACCAAGGCACTTCTG
CAAGACCTGCAGAAGGTACTGGACGCGAGGGGGAGCGTTAAGGAACGTTCCCGTTGGAGG
AGGCTGCAGGCGGAACAAGAGGAGCAAAGGAAGCAGCTCCAAATCTCCCATCACAAGTGG
CAGCAGCGACCAACAAGCCAGCGGCGGCTCCGTGAGCAGCATCACCTCAAGAGGCAGCAG
CGGTCTCGCCGGTGACATACTGAACCGGGCAACAGCGCAAATACCGGGCCTCCGCTACAT
GTCTCCAAGCAGCTTGAACAACCTAGGTGAACTGGGTAATTCGGGAGAGATGGGGTTGAA
CTATGGTGGGATTCCCGGTTCTATGGTTGGGGACGCTGGGAATTTGAGCTTCCAATTGGG
TGGTAACAGCTCGCTCTTATCGATGGGAAGCCTCGAGCACCAGTGGCGTCTGCATCAGGT
CACACAGCAGTTCCCCTTCATGAGTAGCTTGGACAAGCACGCCAATCTGTACTCAATGGA
TGCTTCCACGAGTGGCGTTGAGCCGTCAGGTCTCGCGGGCCATCATCAGGTCAGGCCAAA
TCTGTTCATGAACTCTGGGGCTTCTCACACGAGCTCAGTGAAGGTGGAAAATATCAACCA
AGATCAGTTGAACTTGTCAAGGCAGTTGATGGGGAATTCAGGGGCTAATGACCAGTTCTG
GAGTGTTGGTGGTGGTAGTGGTGGCGGTGGTGTTGGTGGTGGTGGTAGCGCATGGAC
CGATCTTTCGGGTTTTAGCTCTTCTTCCACAAGCAGTGCTTTATAG

>EucaDof10_Eucalyptus_grandis_CDS sequence_forward
ATGGTTTTCTCATCCCTTCCAGCTTATCTTGATCCATCCAACTGGCAACAGCAACAAAAT
CATCACCATCATCAAGCTGGAAATAGTGGTCTCAACAATCAGCTCCCACCTGCTCCTCCG
CCGCCACCGCCTCCTCCTCCACAACCTCATGGAGGAGCTGGAGGCAACCCTGGCGGCTCG
ATCCGTCCGGGCTCGATGGTTGAGCGGGCCCGGCTAGCCAACATACCCATGCCAGAGGCA
GCGCTCAAGTGCCCGCGATGCGATTCCGCCAACACGAAGTTTTGCTACTTCAACAACTAC
AGCCTCTCTCAACCAAGGCACTTCTGCAAGACCTGCAGAAGGTACTGGACGCGAGGGGGA
GCGTTAAGGAACGTTCCCGTTGGAGGAGGCTGCAGGCGGAACAAGAGGAGCAAAGGAAGC
AGCTCCAAATCTCCCATCACAAGTGGCAGCAGCGACCAACAAGCCAGCGGCGGCTCCGTG
AGCAGCATCACCTCAAGAGGCAGCAGCGGTCTCGCCGGTGACATACTGAACCGGGCAACA
GCGCAAATACCGGGCCTCCGCTACATGTCTCCAAGCAGCTTGAACAACCTAGGTGAACTG
GGTAATTCGGGAGAGATGGGGTTGAACTATGGTGGGATTCCCGGTTCTATGGTTGGGGAC
GCTGGGAATTTGAGCTTCCAATTGGGTGGTAACAGCTCGCTCTTATCGATGGGAAGCCTC
GAGCACCAGTGGCGTCTGCATCAGGTCACACAGCAGTTCCCCTTCATGAGTAGCTTGGAC
AAGCACGCCAATCTGTACTCAATGGATGCTTCCACGAGTGGCGTTGAGCCGTCAGGTCTC
GCGGGCCATCATCAGGTCAGGCCAAATCTGTTCATGAACTCTGGGGCTTCTCACACGAGC
TCAGTGAAGGTGGAAAATATCAACCAAGATCAGTTGAACTTGTCAAGGCAGTTGATGGGG
AATTCAGGGGCTAATGACCAGTTCTGGAGTGTTGGTGGTGGTAGTGGTGGCGGTGGTGTT
GGTGGTGGTGGTGGTAGCGCATGGACCGATCTTTCGGGTTTTAGCTCTTCTTCCACAAGC
AGTGCTTTATAG

>EucaDof10_Eucalyptus_grandis_protein sequence
MVFSSLPAYLDPSNWQQQQNHHHHQAGNSGLNNQLPPAPPPPPPPPPQPHGGAGGNPGGS
IRPGSMVERARLANIPMPEAALKCPRCDSANTKFCYFNNYSLSQPRHFCKTCRRYWTRGG
ALRNVPVGGGCRRNKRSKGSSSKSPITSGSSDQQASGGSVSSITSRGSSGLAGDILNRAT

AQIPGLRYMSPSSLNNLGELGNSGEMGLNYGGIPGSMVGDAGNLSFQLGGNSSLLSMGSL
EHQWRLHQVTQQFPFMSSLDKHANLYSMDASTSGVEPSGLAGHHQVRPNLFMNSGASHTS
SVKVENINQDQLNLSRQLMGNSGANDQFWSVGGGSGGGGVGGGGGSAWTDLSGFSSSSTS
SAL+

>EucaDof11_Eucalyptus_grandis_genomic sequence_reverse
ATGCCGGGGGAAGCGAGAGACCCGGCGATCAAGCTCTTCGGGAAGACGATCTCCCTGCCC
GATATTCCTCCTGGGGATGCGGCGGCGGCCGCCGTCGACACTGGCCAGGTTCCGTCCGCG
GATCCTCCCTCGCCCGCCCGCGAGGATATAGGCATGGCCGGCGTCGACGGAGGAGAGGAA
GAGGAAGGGGAGGAGGATGAATACGAAGAAGACGACGACGAGGAAGGCCGCGATGATAAG
GTACGGGTTTTTTTGGCTCCTGTGGTTTTTCGCTCGGTTTGGAGAGTCTTGTTTGGTTTT
TGGCGGATGCCCAAATGCTCAGATTTTGAGAATTTGAATTGGGTCGGGCTTCGTTCTTCT
TGTTGACTTCGGTTGACATGCTTTTGCATTGATCCTGGATCTTCTTTTTCTTCCTTTTTA
TTATTTGTGCGGAAATGAATCTTGAAATGGGGGGGCTTCTTGAATCTGAATGAACTGAGA
TTTGTCTATTCGAATGAGGAGTTGAAACTTGAGTATTCATTATGGAAGTTCGAGGATGTT
CTTGTGGGGGATTTATGGTATGGTAGGTGGATCTTTCCTTGTATTGATGCTTGGAGGCCC
AGTAGAAGCTTAGGTTAATTTGCTGGTGCCTGGTGCTTTTGCTTTTACGCGAGTAAGGAT
CACCTGAAACTAGTTGGATGCTCTGCGCAAGTTCATGTTCGTGAGCTGGCTCTGATCTCC
ACGGGTCCAAATTCCAATCTTCTCTGTCGAGGCCAATTCTAATTCTCACACAAACGAACT
GGAGACTTTTGGAGGTTCCGAAGTCGACATAATCAAGCTCTGGTCGTTTCTCTTCCTCCT
TTTATTCCTTGACAAGACGAATCACGGCCGGTTCTTCTGTTCAAAACACCCGAAAAGCGT
CTGTCGTGATTGCTCTTAATAGCCTGGGTACTGGGTGCTGCTACACGTTGGATCAATGCT
CTTATCTTTTTCTTGATCTTTTCCTAAAAGTCGCTTGTCCGTGGCTTTTTGACTCCATAG
GCGAAGACCTCAATTCCGTTTATCTTTCTATTCCGTAGGACAACCCATCAAGTCCAGTCT
GCCAAAGAAGTCCGCCATTTACCTGTTGCATATGCAATTTCTAAAGGAATTGCAGAAAGA
TAAGGACATTTATAAATCTGAAAACCACCTGACTGGATTTCGTTTGTTGGCACTATTCAA
AATCGCTTAGAATATCTGATTTGCTAGTTTTTGCTTAATTTAGCAATAGCCAATCTTACA
TATAGTTTGGTCATTTGCTGATTGAATTTAAGTGCTTGTATGGCTATTCCTTCAGTTGGA
CTTCCTTTCGATATAAGTTTTGTCCGTAATTGCTATCATACCATTCGTGGCCTGTTCCTC
TGTCGCTGTGGAGGTGGAGTGTTTAGTTTGAACCTCAGAAAATTGATTAGGGTATGTGAA
AATATGTTTTTCTCTCTATCCAGGAAACACAGGATGAAAAACCAGCTATAGCCAAGCTGG
AAAATGGAGCCCAATTCATGAATCTCGAGGAGGGCACAAACCGGGACGCAATATCGGTGA
CAAATGAGAACCCTAGGACGCCTTCAGTCGACAAGGAGAGCACTACTCCAAGGACGTCGA
AATCAGAAGAGGAGCAGAGCGACACGAGCAACTCGCAAGAGAAGGTGCTCAAGAAACCTG
ACAAGATACTTCCTTGCCCTCGATGTAATAGCATGGACACCAAATTCTGTTACTACAACA
ACTATAATGTGAACCAGCCCCGACACTTCTGCAAGAACTGCCAGAGATACTGGACAGCTG
GTGGAACCATGAGGAATGTTCCTGTGGGTGCTGGCCGCCGCAAGAACAAGAACTCGGCAT
CTCATTACCGTCATCTAGCCGTTCCTGAAGCCCTGCAAAGCGTACGGACGGACTTCCCTA
ATGGGGTCCATCACCCGCCATTGAAACCTAATGGCACTGTCCTCACCTTCGGATCAGACG
CACCACTATGTGAATCCATGGCTTCGGTTTTGCATTTGGCTGATAAGGCAATGCATAACA
GCAGGAAAAATGGATTCCATAAACCTGAAGGAGTGATAATGCCAAGCCCTTGTGGAAGTG
TGAGAAGTGGGGTTGAGACTATGAATGGATCTTCTGCCACCGCAGCGAATTCCAACCTTG
TAGAGAGTAAAAATGGAACCAAAGAAATTGTGATGCCGAATTGTCAGGGTTTTCCTCCTC
AGCTACCGTGCTTTCCGGGGGCTCCTTGGCCTTATGCATGGAATTCTGCTCCATGGAGCC
CGCCAATATCCCCTGGCCCGCAAACTTTCTGTCCACCTGGTTTTGCTGTGCCGTTTTACG
CAGGGCCAGCTTATTGGAGTTGCAGTGTGCCAGGTGCGTGGAACATTCCTTGGGTTGCCC
AGCCAACTTCTCCGAACCACAGCGCCACGAGCTCCGCACCTAATTCTCCGACCTTGGGGA
AACACTCGAGGGATGAGAAAATCGGGAAACCAAACGACTCCACTGAAGAGGAGCCAAAGG
AGGCTAATCCTGAGAGATCCTTGTGGATTCCAAAGACATTAAGGATTGATGATCCAGGAG
AAGCTGCAAAGAGCTCGATCTGGGCGACACTTGGGATTAAAAACGAAGCTTCTAATTCAG
CTGGCAGGGGAGTACTCTTAAAGGCCTTCCAGTCTAAGAGCGAGGAAAAGAACCACGTGG
TCGAAGCTTCTTCTGTGTTACAAGCAAATCCAGCCGCATTGTCTCGATCACGCAACTTCC
AAGAGAGCTCGTAA

>EucaDof11_Eucalyptus_grandis_CDS sequence_reverse
ATGCCGGGGGAAGCGAGAGACCCGGCGATCAAGCTCTTCGGGAAGACGATCTCCCTGCCC
GATATTCCTCCTGGGGATGCGGCGGCGGCCGCCGTCGACACTGGCCAGGTTCCGTCCGCG
GATCCTCCCTCGCCCGCCCGCGAGGATATAGGCATGGCCGGCGTCGACGGAGGAGAGGAA
GAGGAAGGGGAGGAGGATGAATACGAAGAAGACGACGACGAGGAAGGCCGCGATGATAAG
GAAACACAGGATGAAAAACCAGCTATAGCCAAGCTGGAAAATGGAGCCCAATTCATGAAT
CTCGAGGAGGGCACAAACCGGGACGCAATATCGGTGACAAATGAGAACCCTAGGACGCCT

TCAGTCGACAAGGAGAGCACTACTCCAAGGACGTCGAAATCAGAAGAGGAGCAGAGCGAC
ACGAGCAACTCGCAAGAGAAGGTGCTCAAGAAACCTGACAAGATACTTCCTTGCCCTCGA
TGTAATAGCATGGACACCAAATTCTGTTACTACAACAACTATAATGTGAACCAGCCCCGA
CACTTCTGCAAGAACTGCCAGAGATACTGGACAGCTGGTGGAACCATGAGGAATGTTCCT
GTGGGTGCTGGCCGCCGCAAGAACAAGAACTCGGCATCTCATTACCGTCATCTAGCCGTT
CCTGAAGCCCTGCAAAGCGTACGGACGGACTTCCCTAATGGGGTCCATCACCCGCCATTG
AAACCTAATGGCACTGTCCTCACCTTCGGATCAGACGCACCACTATGTGAATCCATGGCT
TCGGTTTTGCATTTGGCTGATAAGGCAATGCATAACAGCAGGAAAAATGGATTCCATAAA
CCTGAAGGAGTGATAATGCCAAGCCCTTGTGGAAGTGTGAGAAGTGGGGTTGAGACTATG
AATGGATCTTCTGCCACCGCAGCGAATTCCAACCTTGTAGAGAGTAAAAATGGAACCAAA
GAAATTGTGATGCCGAATTGTCAGGGTTTTCCTCCTCAGCTACCGTGCTTTCCGGGGGCT
CCTTGGCCTTATGCATGGAATTCTGCTCCATGGAGCCCGCCAATATCCCCTGGCCCGCAA
ACTTTCTGTCCACCTGGTTTTGCTGTGCCGTTTTACGCAGGGCCAGCTTATTGGAGTTGC
AGTGTGCCAGGTGCGTGGAACATTCCTTGGGTTGCCCAGCCAACTTCTCCGAACCACAGC
GCCACGAGCTCCGCACCTAATTCTCCGACCTTGGGGAAACACTCGAGGGATGAGAAAATC
GGGAAACCAAACGACTCCACTGAAGAGGAGCCAAAGGAGGCTAATCCTGAGAGATCCTTG
TGGATTCCAAAGACATTAAGGATTGATGATCCAGGAGAAGCTGCAAAGAGCTCGATCTGG
GCGACACTTGGGATTAAAAACGAAGCTTCTAATTCAGCTGGCAGGGGAGTACTCTTAAAG
GCCTTCCAGTCTAAGAGCGAGGAAAAGAACCACGTGGTCGAAGCTTCTTCTGTGTTACAA
GCAAATCCAGCCGCATTGTCTCGATCACGCAACTTCCAAGAGAGCTCGTAA

>EucaDof11_Eucalyptus_grandis_protein sequence
MPGEARDPAIKLFGKTISLPDIPPGDAAAAAVDTGQVPSADPPSPAREDIGMAGVDGGEE
EEGEEDEYEEDDDEEGRDDKETQDEKPAIAKLENGAQFMNLEEGTNRDAISVTNENPRTP
SVDKESTTPRTSKSEEEQSDTSNSQEKVLKKPDKILPCPRCNSMDTKFCYYNNYNVNQPR
HFCKNCQRYWTAGGTMRNVPVGAGRRKNKNSASHYRHLAVPEALQSVRTDFPNGVHHPPL
KPNGTVLTFGSDAPLCESMASVLHLADKAMHNSRKNGFHKPEGVIMPSPCGSVRSGVETM
NGSSATAANSNLVESKNGTKEIVMPNCQGFPPQLPCFPGAPWPYAWNSAPWSPPISPGPQ
TFCPPGFAVPFYAGPAYWSCSVPGAWNIPWVAQPTSPNHSATSSAPNSPTLGKHSRDEKI
GKPNDSTEEEPKEANPERSLWIPKTLRIDDPGEAAKSSIWATLGIKNEASNSAGRGVLLK
AFQSKSEEKNHVVEASSVLQANPAALSRSRNFQESS*

>EucaDof12_Eucalyptus_grandis_genomic sequence_forward
ATGGATCCTTCAAGTGGAACAAACCAGGTATTCTAGGAAGCAAGAAACATTTCTGAGCTG
CTCTTGTTTATCTCATCTCAAGTCAATCTTTCTTTGTTTCTTTCTTCTCCTCTTTTCTTT
TCAGATGGGTTGTGAGATCTAAGCTTTTTAAGTAAAGAGAGAGTAGTGTGATCAAGAAAA
ATTGGATAGGTTAAAGTACTTTTTTGTCTGCCACCACATAGTTCAGTCATAAATCCATGT
TTTCTCAACACTGTTTTTTGTCACCTTAAAATAACAGAGTTGTGTGCAAAAATCACCTCC
TTTGAATCTGATATTGATCAAGATGCTTCATTTGGATGAAAAGTGAGGAAGCAAGACAAA
GGCAAAGATGAAGACTCCAACTTTAGAAATTTGGACAGCTGAGAAAAATATGGACTAGAG
AGTTGCTCACCACCGTGAGAAAACTATCCAGCCCATTTCTTCTCTACGTTTTAAATATTA
TAAGCAAGAAAAATATAGATGGGTGTTTTTATCCTTTTCACGCACAACTGAGAAAATCCT
TGACTTCATCCATCTGATTTGTTGTGTTTTGCGGTCATGCAGGAAATGGGAGCTCATTCA
TTGGAAAGCATGTTGATATCCCCTAAACCACAGCAAGATCAAAGGAAGCCAAGGCCTCAG
CCAGAAGAGGCTCTCCACTGCCCGAGGTGCGAGTCCACCAACACCAAGTTCTGTTACTAC
AACAACTACAGCCTCTTGCAGCCTAGGTACTTCTGCAAGTCCTGTCGAAGGTATTGGACC
AAAGGAGGGACCCTGAGGAATGTTCCTGTGGGTGGAGGGTGCAGGAAGAACCGTAGGTCC
TCGTCCTCATCATCATCCAGAAGAGCCCAAGACCAGGGAGGACCAATCTCAACTTTGATC
AACAACAATGATTACGACACCCCAAACCCATTTACTACCTTGCACTCTCTAACTTATGAT
CAGTCCAGTGATCATATGTCCCTTCCATTCCTTGGGCTCCACAAGTTCCCATACCAGCTT
GGGCCTCATGAGCTGAGCCATGATATCCTCTCCTATGCCAACAGATCAAAGAACGGCCAT
GGTGACATGAACGGAGGCTTTCTTGATTCACTGAAGAACGAGTTTGGCTTGCTTGAGAAC
ACCCAGAATAATGTTCATGACAGTCATCATCGTCAGAATCTGTATTATGGGTATGGATCA
GATGGGATGCACAATAACATGGGAGATCCAGTGGACAGTGGTGGTGCTCCATCATCATGT
GATCATGAAATGATGAGCCCTGCTGGTTCTGCAAATGAAGCTCTTGGAGCGACATCAATG
AAGCAAGAGCTGTTGTGTTCTGATAGAGACGGCGAAAATAGGGTTTTTTGGGGTTGTCCG
TGGCAGATGAATGGAGATCAATATCGCGGGAATGGTGTAATTGGTGATGTTAATTCAGGG
AGGGGAAGTTGGAATGGTCTTGGTTCATCTTGGTATGGGCTTCTTAATAGCCCTCTCGTG
TAGTGTCGTGTCCTACAGTTTCAAATGTTCAAACCTGGAAAAGAATGGCAGTAGATTAGA
ATGTTTTGTTCTTGCTGTGCATGATCAGGCCCAACTGAATCATCTCGGGAGTTCTTCAGT
TTTATCTTTTCTTACATCTTTTGATTCCTCCTTAAAATGTGGTACTGAAGTGGTGAGTTT

GTCTGAAATCGGTCTGGAAGGACCATGTCCTGTCATTCTTCAGTGAAGAAAAGAGAATTG
AGTAGTCTTCAGAATGATTTTACACGGTCTGTTGTGCAGTCTGAGTACTCTGAGACTTCC
ATTGATTAG

>EucaDof12_Eucalyptus_grandis_CDS sequence_forward
ATGGATCCTTCAAGTGGAACAAACCAGGAAATGGGAGCTCATTCATTGGAAAGCATGTTG
ATATCCCCTAAACCACAGCAAGATCAAAGGAAGCCAAGGCCTCAGCCAGAAGAGGCTCTC
CACTGCCCGAGGTGCGAGTCCACCAACACCAAGTTCTGTTACTACAACAACTACAGCCTC
TTGCAGCCTAGGTACTTCTGCAAGTCCTGTCGAAGGTATTGGACCAAAGGAGGGACCCTG
AGGAATGTTCCTGTGGGTGGAGGGTGCAGGAAGAACCGTAGGTCCTCGTCCTCATCATCA
TCCAGAAGAGCCCAAGACCAGGGAGGACCAATCTCAACTTTGATCAACAACAATGATTAC
GACACCCCAAACCCATTTACTACCTTGCACTCTCTAACTTATGATCAGTCCAGTGATCAT
ATGTCCCTTCCATTCCTTGGGCTCCACAAGTTCCCATACCAGCTTGGGCCTCATGAGCTG
AGCCATGATATCCTCTCCTATGCCAACAGATCAAAGAACGGCCATGGTGACATGAACGGA
GGCTTTCTTGATTCACTGAAGAACGAGTTTGGCTTGCTTGAGAACACCCAGAATAATGTT
CATGACAGTCATCATCGTCAGAATCTGTATTATGGGTATGGATCAGATGGGATGCACAAT
AACATGGGAGATCCAGTGGACAGTGGTGGTGCTCCATCATCATGTGATCATGAAATGATG
AGCCCTGCTGGTTCTGCAAATGAAGCTCTTGGAGCGACATCAATGAAGCAAGAGCTGTTG
TGTTCTGATAGAGACGGCGAAAATAGGGTTTTTTGGGGTTGTCCGTGGCAGATGAATGGA
GATCAATATCGCGGGAATGGTGTAATTGGTGATGTTAATTCAGGGAGGGGAAGTTGGAAT
GGTCTTGTGAAGAAAAGAGAATTGAGTAGTCTTCAGAATGATTTTACACGGTCTGTTGTG
CAGTCTGAGTACTCTGAGACTTCCATTGATTAG

>EucaDof12_Eucalyptus_grandis_protein sequence
MDPSSGTNQEMGAHSLESMLISPKPQQDQRKPRPQPEEALHCPRCESTNTKFCYYNNYSL
LQPRYFCKSCRRYWTKGGTLRNVPVGGGCRKNRRSSSSSSSRRAQDQGGPISTLINNNDY
DTPNPFTTLHSLTYDQSSDHMSLPFLGLHKFPYQLGPHELSHDILSYANRSKNGHGDMNG
GFLDSLKNEFGLLENTQNNVHDSHHRQNLYYGYGSDGMHNNMGDPVDSGGAPSSCDHEMM
SPAGSANEALGATSMKQELLCSDRDGENRVFWGCPWQMNGDQYRGNGVIGDVNSGRGSWN
GLKKRELSSLQNDFTRSVVQSEYSETSID+

>EucaDof13_Eucalyptus_grandis_genomic sequence_reverse
ATGGAGCAAGAGAGCAAGCCACCGCCGCTGGATCGGAGGCCGACATCGAGGCCGGCCCAA
CCGGACGGACAGCAACAGCCGCCGCCGCCGCACCACCAGCCCCAGAAGTGCCCTCGCTGC
GACTCCATGAACACCAAGTTCTGCTACTACAACAACTACAGCCTCTCCCAGCCCCGCTAC
TTCTGCAAGACCTGCCGCCGCTACTGGACCCACGGCGGAACCCTAAGGAACGTCCCCGTC
GGCGGCGGCTGCCGGAAGGCCAAGCGCTCCAAGTCCTCCTCCTCCTCCTCCTCCACATCC
GCCTCCGCCTCCTCCGGCGACAGCCTCACCCCGCACTCATCCCTGCAGGTCCTGCCTCAA
AGCCTCAGGGATGTTGCAGCTGGGAACCTGGCCCCACCGCCGGCCGCGGCGGCCGGGGAG
GGGCCTCTGGGTTACTCGTTCTATCCGGGAGGTGGGTTCTTGTCCTCTTTGGCGGCAGTT
CAGTCTATCAACCAGAGCCAGCCCTTCAACATTGGGGGTGCTACAACAAGTAACAACAAC
AGCAGCAACAATTTGGGCTTTTTCCAAGGGTTTGGAGGTTTGGCTTCCTTAAGCAGCTCT
CACCACCACCAGCAGATTCATCAACAAGCACAGTTTCTTCAGAGCGATTGGAGTCACGGC
TTCGGCATGGGCAGTAATGCTAATACCGGTAGTAGTAGCAGTAACAATAATAGCCCTACT
ATTTCACATGTACATCATCAGCATGGTAACTTGTGGAGTGGAAGCACGAGCGCTGGTATT
AGCGGTCACGGTGGCGGCGAGAGTAATAGAATCGGTCTGAATGTTAACCGCCCATCCGAT
CATCACCACTTGCCCCATCTTCCAGGATTTGGCCCTCCTCCATAA

>EucaDof13_Eucalyptus_grandis_CDS sequence_reverse
ATGGAGCAAGAGAGCAAGCCACCGCCGCTGGATCGGAGGCCGACATCGAGGCCGGCCCAA
CCGGACGGACAGCAACAGCCGCCGCCGCCGCACCACCAGCCCCAGAAGTGCCCTCGCTGC
GACTCCATGAACACCAAGTTCTGCTACTACAACAACTACAGCCTCTCCCAGCCCCGCTAC
TTCTGCAAGACCTGCCGCCGCTACTGGACCCACGGCGGAACCCTAAGGAACGTCCCCGTC
GGCGGCGGCTGCCGGAAGGCCAAGCGCTCCAAGTCCTCCTCCTCCTCCTCCTCCACATCC
GCCTCCGCCTCCTCCGGCGACAGCCTCACCCCGCACTCATCCCTGCAGGTCCTGCCTCAA
AGCCTCAGGGATGTTGCAGCTGGGAACCTGGCCCCACCGCCGGCCGCGGCGGCCGGGGAG
GGGCCTCTGGGTTACTCGTTCTATCCGGGAGGTGGGTTCTTGTCCTCTTTGGCGGCAGTT
CAGTCTATCAACCAGAGCCAGCCCTTCAACATTGGGGGTGCTACAACAAGTAACAACAAC
AGCAGCAACAATTTGGGCTTTTTCCAAGGGTTTGGAGGTTTGGCTTCCTTAAGCAGCTCT
CACCACCACCAGCAGATTCATCAACAAGCACAGTTTCTTCAGAGCGATTGGAGTCACGGC
TTCGGCATGGGCAGTAATGCTAATACCGGTAGTAGTAGCAGTAACAATAATAGCCCTACT

ATTTCACATGTACATCATCAGCATGGTAACTTGTGGAGTGGAAGCACGAGCGCTGGTATT
AGCGGTCACGGTGGCGGCGAGAGTAATAGAATCGGTCTGAATGTTAACCGCCCATCCGAT
CATCACCACTTGCCCCATCTTCCAGGATTTGGCCCTCCTCCATAA

>EucaDof13_Eucalyptus_grandis_protein sequence
MEQESKPPPLDRRPTSRPAQPDGQQQPPPPHHQPQKCPRCDSMNTKFCYYNNYSLSQPRY
FCKTCRRYWTHGGTLRNVPVGGGCRKAKRSKSSSSSSSTSASASSGDSLTPHSSLQVLPQ
SLRDVAAGNLAPPPAAAAGEGPLGYSFYPGGGFLSSLAAVQSINQSQPFNIGGATTSNNN
SSNNLGFFQGFGGLASLSSSHHHQQIHQQAQFLQSDWSHGFGMGSNANTGSSSSNNNSPT
ISHVHHQHGNLWSGSTSAGISGHGGGESNRIGLNVNRPSDHHHLPHLPGFGPPP#

>EucaDof14_Eucalyptus_grandis_genomic sequence_forward
ATGCAAGAGACGACAGCGTTTCAGAGCACGAAGCTACTGTTCCCAGAGCAAGAGAGGCTG
ATATGCCCACGTTGCGATTCGACCAACACCAAGTTCTGCTACTTCAACAACTACAACCTC
TCCCAGCCCCGCCATTTCTGCAAGAACTGCAAGCGCTACTGGACCAAAGGCGGCTCTCTC
CGCAACATCCCCGTCGGCGGCGGCACCCGCAAGAAATCCTCCACCAAGCGGGCCTCCAAC
CCGAAACGCCAGAACCCGGATCCGGATCCTGACCCGAGACCCAGTAAGCACGGGCTGCAG
CTCCAAGAGTCGGCCACAGCCGACTCCTTGCAGTCTAACGCCACAGAAGGGGCCCTGTTG
GAGGGTCCGAACTCGAGCGGATTGGGGGCGGGATCCGGTGGATATGGGGTTAGCCTGAAC
CCGGGCTCGAGCTTGGGTTTGGGTTCTGTCATGGGGCCTGAACCAGAGCTTGACTCGCGA
CGTGGCGGCGGAGGGCGGGAGGATCATTTTGGACATGCAAGATGGTCTTGCTATCTAGAG
TACACCAGCTTCAAGACTACTACGGAGGAGAGATGGTGA

>EucaDof14_Eucalyptus_grandis_CDS sequence_forward
ATGCAAGAGACGACAGCGTTTCAGAGCACGAAGCTACTGTTCCCAGAGCAAGAGAGGCTG
ATATGCCCACGTTGCGATTCGACCAACACCAAGTTCTGCTACTTCAACAACTACAACCTC
TCCCAGCCCCGCCATTTCTGCAAGAACTGCAAGCGCTACTGGACCAAAGGCGGCTCTCTC
CGCAACATCCCCGTCGGCGGCGGCACCCGCAAGAAATCCTCCACCAAGCGGGCCTCCAAC
CCGAAACGCCAGAACCCGGATCCGGATCCTGACCCGAGACCCAGTAAGCACGGGCTGCAG
CTCCAAGAGTCGGCCACAGCCGACTCCTTGCAGTCTAACGCCACAGAAGGGGCCCTGTTG
GAGGGTCCGAACTCGAGCGGATTGGGGGCGGGATCCGGTGGATATGGGGTTAGCCTGAAC
CCGGGCTCGAGCTTGGGTTTGGGTTCTGTCATGGGGCCTGAACCAGAGCTTGACTCGCGA
CGTGGCGGCGGAGGGCGGGAGGATCATTTTGGACATGCAAGATGGTCTTGCTATCTAGAG
TACACCAGCTTCAAGACTACTACGGAGGAGAGATGGTGA

>EucaDof14_Eucalyptus_grandis_protein sequence
MQETTAFQSTKLLFPEQERLICPRCDSTNTKFCYFNNYNLSQPRHFCKNCKRYWTKGGSL
RNIPVGGGTRKKSSTKRASNPKRQNPDPDPDPRPSKHGLQLQESATADSLQSNATEGALL
EGPNSSGLGAGSGGYGVSLNPGSSLGLGSVMGPEPELDSRRGGGGREDHFGHARWSCYLE
YTSFKTTTEERW*

>EucaDof15_Eucalyptus_grandis_genomic sequence_forward
ATGGATTTTCCATCTCTCCCACCTTATCTTGATCCATCCAACTGGCAACAGGTGAGGATT
TTCTTAAACGCTCCATGAATTATCTTTCTTTCATGCTTTCTCAGCCCAAAAAGCAAGAAA
AAACTTCAAAAAGAAAAGATATTCTCTCCACTTTTGCTGGTCTTCTTTGCTGAATATACC
AGACATTTGATCTTTGCTTTCTTCTTCTTCTTCTTCTTCTTCAGCAACAGACAGATC
ATCACCAAACTGGAAGCTCCAGCACAAATACTCAGCTCCTGCCGCCACCACCACAACCTA
ATGGAGGCGGTGGTGCCAGCTCGATCAGGCCCCGTTTGATGGCTGACGAGGCCCGCCTGG
CCAACATAACGATGCCAGAGGGGCCGCTGAAATGCCCTCGATGTGAATCGACCAACACAA
AGTTCTGCTACTTCAATAACTACAGCCTCTCCCAGCCGAGACACTTCTGCAAAACGTGCA
GAAGGTACTGGACGCGAGGAGGAGCGTTAAGAAATGTCCCTGTTGGAGGCGGTTGCCGGA
GAAACAACAAGAGGAGCAAATCGGGGAGCTCGAAATCAGTGGCAAGTAGCAGCCATCGGC
AAATCACAGCTCCGAGATCGGGCTCTACTTGCATACCGACCTCAAATGGAGGGGCCAATG
TTGACACCACTGAAATATTAGGGGTCGGGACGCAGCTCCCACCGCTCGGCCACTTCATGG
CTCCTATGAACTATCTTCCTGATAACTTGACTTTATCAGCTGGCGCTGGACTTAATTATA
ATACTGGAGTGCCTTCACCACTAGGAGACTTGAATTTCCTGGCCGCGACCACCTGCGATG
GTGGAGCTTTAGAGCGGTGGCGCTTGGAGCAAGCACAGCAGCTTTCCTCGTCGGGCAGGC
TGGATCCAAACCTGGGATTGTACCCGTCTGAAGGCAGGTTTGAGTCATTTAGAGCTCTTA
ATGTGAGGCCTAATGCACCGCCTTCATCAGCTGGGCTTACTGCTCAGCTCTCCTCATCGG
TGAAAATGGAAGACAACAATCAGCCTTTCTCGAGGCAGTTTTACGGGATTAACCATGAAA

ATCATCCCTTCTTGAGCAGTGCCAGCGCATGGATGGATCTTTCTGGTTTCAGCTCTTCTT
CCACTCGCAACCCCTTATAGAGATCTCTTGGTTAACTTGATCCTCTCTTTTTGAAGCTTT
CATAGTTCTCCTCCTTTCTTTCAAATTTTTGGTCATCTCAACTGATCATTCAAAAAAAGA
ATGAAACCCTAGAAATATTTAGTGGAGATGGTTGCTGACTGCTCTGCTCTTAAAGATGAA
ACCCTAGAAATATCTCTCACTTTGGTGTTTTTCTTTTCTTTTTTTTCCTTTCGGTGTTCA
TCTGTTAGGTTTGAACTGTCACCTGCTCACCATAGAGACTCTGTATGCCTTTGCCAAGTT
GTGATCATATTATGTTCCTGTCGAATGCATAATTCACTGGCTGATCTCCTTAATCAAGGT
TTCTTTCGTTCTTTAATTTTGTGACGATTGGTTGAGGCAATGTCTTCAGTTACCATTTGT
TTACCGTGAGTAGATGTTATTTGGGAGTTATAAATTATACGTTTTCTTATTTTGAGTAAC
ACGAGAGTGCCAACGTTAGAATCTGCAGCTCAACAGTTGTGTGTAGCATACTTTCAGAAA
TTATTTTGACAAGTGAAGTAGAAGATTGCATCTCAAAGCAACGTGCATGAATTGATGGTC
ATATGCAAAACACTTGTGGTTTTGAGTACGGACCAGCAATTTGAAAGTATATAGGAGCTG
ATTCAAGAATTTAAATCCTCATTGCCTATAAGGCTCCTTATAGTACAAGGACTCCATCTT
TATACTGATTTTAGCCTCTTACTTCTTCCATTGTGCCTTTGCTCCTCGAGCCTACATGAT
TCGTAACATAACTCATTTATATCTTGCTTTTCACTAGTAAAATTTGAATTAAATTTAAGG
GCGTGCTCATGTGTGTATGGACTTATCTTGTGAAGGGATCTAAACTTTTAGAACAAATTT
CCGTCTAACAATCCCAATCTGTCGAACTTTCGGCATAATATTTCTGGGTCATAGCAACTA
GTTAACGTCTAACTCTTGCCGCCTCTTAAATTTATCGACAACCGACCAAGTTTCTATCCA
ATTACATATGCGATTACTAATAACTTAACTAATTTACTATCAGCTTGAGAAGCCACTTTT
TGGGGAAACATTGAGCACCAGACGCCGTGAATATTTTTGCTGGTCTTTCATGGATACCTT
TTGTTTAGTGAAAATTTCTCACTGTAATCCCTTCGGCCGCAAAGGGAATATATTCTGGTT
ATATATGCTCGAGGCATCCGAGCATATCTTTTGATAGCCAAACCTGAAAATAGAAACAAT
AAGTACCGTGAAGTTGTTGAGCCAGGGGCTAATTTAATCAACGGACATTTATATGTGTGG
ATTAGGCTTGTTCTGCGTCTTACTTGGGTCGTGATGGTTAATATTACTACTTGA

>EucaDof15_Eucalyptus_grandis_CDS sequence_forward
ATGGATTTTCCATCTCTCCCACCTTATCTTGATCCATCCAACTGGCAACAGCAACAGACA
GATCATCACCAAACTGGAAGCTCCAGCACAAATACTCAGCTCCTGCCGCCACCACCACAA
CCTAATGGAGGCGGTGGTGCCAGCTCGATCAGGCCCCGTTTGATGGCTGACGAGGCCCGC
CTGGCCAACATAACGATGCCAGAGGGGCCGCTGAAATGCCCTCGATGTGAATCGACCAAC
ACAAAGTTCTGCTACTTCAATAACTACAGCCTCTCCCAGCCGAGACACTTCTGCAAAACG
TGCAGAAGGTACTGGACGCGAGGAGGAGCGTTAAGAAATGTCCCTGTTGGAGGCGGTTGC
CGGAGAAACAACAAGAGGAGCAAATCGGGGAGCTCGAAATCAGTGGCAAGTAGCAGCCAT
CGGCAAATCACAGCTCCGAGATCGGGCTCTACTTGCATACCGACCTCAAATGGAGGGGCC
AATGTTGACACCACTGAAATATTAGGGGTCGGGACGCAGCTCCCACCGCTCGGCCACTTC
ATGGCTCCTATGAACTATCTTCCTGATAACTTGACTTTATCAGCTGGCGCTGGACTTAAT
TATAATACTGGAGTGCCTTCACCACTAGGAGACTTGAATTTCCTGGCCGCGACCACCTGC
GATGGTGGAGCTTTAGAGCGGTGGCGCTTGGAGCAAGCACAGCAGCTTTCCTCGTCGGGC
AGGCTGGATCCAAACCTGGGATTGTACCCGTCTGAAGGCAGGTTTGAGTCATTTAGAGCT
CTTAATGTGAGGCCTAATGCACCGCCTTCATCAGCTGGGCTTACTGCTCAGCTCTCCTCA
TCGGTGAAAATGGAAGACAACAATCAGCCTTTCTCGAGGCAGTTTTACGGGATTAACCAT
GAAAATCATCCCTTCTTGAGCAGTGCCAGCGCATGGATGGATCTTTCTGGTTTCAGCTCT
TCTTCCACTCGCAACCCCTTATAGAGATCTCTTGGTTAACTTGATCCTCTCTTTTTGAAG
CTTTCATAGTTCTCCTCCTTTCTTTCAAATTTTTGGTCATCTCAACTGATCATTCAAAAA
AAGAATGAAACCCTAGAAATATTTAGTGGAGATGGTTGCTGACTGCTCTGCTCTTAAAGA
TGAAACCCTAGAAATATCTCTCACTTTGGTGTTTTTCTTTTCTTTTTTTTCCTTTCGGTG
TTCATCTGTTAGGTTTGAACTGTCACCTGCTCACCATAGAGACTCTGTATGCCTTTGCCA
AGTTGTGATCATATTATGTTCCTGTCGAATGCATAATTCACTGGCTGATCTCCTTAATCA
AGGTTTCTTTCGTTCTTTAATTTTGTGACGATTGGTTGAGGCAATGTCTTCAGTTACCAT
TTGTTTACCGTGAGTAGATGTTATTTGGGAGTTATAAATTATACGTTTTCTTATTTTGAG
TAACACGAGAGTGCCAACGTTAGAATCTGCAGCTCAACAGTTGTGTGTAGCATACTTTCA
GAAATTATTTTGACAAGTGAAGTAGAAGATTGCATCTCAAAGCAACGTGCATGAATTGAT
GGTCATATGCAAAACACTTGTGGTTTTGAGTACGGACCAGCAATTTGAAAGTATATAGGA
GCTGATTCAAGAATTTAAATCCTCATTGCCTATAAGGCTCCTTATAGTACAAGGACTCCA
TCTTTATACTGATTTTAGCCTCTTACTTCTTCCATTGTGCCTTTGCTCCTCGAGCCTACA
TGATTCGTAACATAACTCATTTATATCTTGCTTTTCACTAGTAAAATTTGAATTAAATTT
AAGGGCGTGCTCATGTGTGTATGGACTTATCTTGTGAAGGGATCTAAACTTTTAGAACAA
ATTTCCGTCTAACAATCCCAATCTGTCGAACTTTCGGCATAATATTTCTGGGTCATAGCA
ACTAGTTAACGTCTAACTCTTGCCGCCTCTTAAATTTATCGACAACCGACCAAGTTTCTA
TCCAATTACATATGCGATTACTAATAACTTAACTAATTTACTATCAGCTTGAGAAGCCAC
TTTTTGGGGAAACATTGAGCACCAGACGCCGTGAATATTTTTGCTGGTCTTTCATGGATA

CCTTTTGTTTAGTGAAAATTTCTCACTGTAATCCCTTCGGCCGCAAAGGGAATATATTCT
GGTTATATATGCTCGAGGCATCCGAGCATATCTTTTGATAGCCAAACCTGAAAATAGAAA
CAATAAGTACCGTGAAGTTGTTGAGCCAGGGGCTAATTTAATCAACGGACATTTATATGT
GTGGATTAGGCTTGTTCTGCGTCTTACTTGGGTCGTGATGGTTAATATTACTACTTGA

>EucaDof15_Eucalyptus_grandis_protein sequence
MDFPSLPPYLDPSNWQQQQTDHHQTGSSSTNTQLLPPPPQPNGGGGASSIRPRLMADEAR
LANITMPEGPLKCPRCESTNTKFCYFNNYSLSQPRHFCKTCRRYWTRGGALRNVPVGGGC
RRNNKRSKSGSSKSVASSSHRQITAPRSGSTCIPTSNGGANVDTTEILGVGTQLPPLGHF
MAPMNYLPDNLTLSAGAGLNYNTGVPSPLGDLNFLAATTCDGGALERWRLEQAQQLSSSG
RLDPNLGLYPSEGRFESFRALNVRPNAPPSSAGLTAQLSSSVKMEDNNQPFSRQFYGINH
ENHPFLSSASAWMDLSGIRAYLLIAKPENRNNKYREVVEPGANLINGHLYVWIRLVLRLT
WVVMVNITT*

>EucaDof16_Eucalyptus_grandis_genomic sequence_reverse
ATGACACAGGCGTTTTTGCTGACACGCGCATTCCACATTCTCTTGTCCTTTGATTTTTCT
CAGGGCACTCTTCACGACGACTCGGGAATGGACACGTCCTCGCCTAACTCAGGCGACATG
CTCGCGTGCTCCAGGCCCTTGATGGAGAGGAGGCTCCGGCCGCAGCACGACCAGCCCTTG
AAGTGCCCCAGGTGCGAATCCACTCACACCAAGTTCTGCTACTACAACAACTACAGCCTT
TCGCAGCCCCGGTACTTCTGCAAGACCTGCCGCCGCTACTGGACCAAAGGCGGCACCCTC
CGCAACATCCCCGTCGGAGGCGGCTGCCGGAAGAACAAGAAAGCCTCTAATTCCGGCAGT
AATAGCAACAACAGTGCCAGCAGCAAGAAATTGGCCAACAATGATCATCAAGGATATGGG
GTCCCTAGCCACCCCGGGCAGTCGTCAGCGTCCTCGTCTCGTGGCGACCACTCGACCGAC
CTCCACCTCTCGTTCTCGCCTGAGGTGCAGCTTTCGCACCTCCACGGCTTGCTCGGCGTC
CATGGGGCCCCCTTCGCGGGCCCGGGCAACAATTTCTCGGGGCTCGAGAACCCGAGGCCC
ATCGATTTCTTGGAGGGCAAGGTAATGAGTTACGACAGCAACTTCATGCTGAGCGGCGAT
AGCCCTCATCCTCACATGGGCGTGGTGGGAGATCTCGGTGGCTCGACTGGGTTGGCCCCG
GGCTTCCATGGCTTCTGCGGTCCATACGGGATGGCCATCGACGGTGGTCATAGCGGAAAT
GCTAACAGTTTCATGGAGAATTGCCAGAGATTGAACATGCTCGTTCCATACGATCAGGGC
CATGATCAGGAGCAGATCAACGCGATGGATGTGAAGCCGAACCCGAAGCTCTTGTCGCTG
GAGTGGCAAGACCAAGGCTGCTCCGACGCGAGCAAGAACGGGTATGGCTATTTCAACGGT
CTAGGACCGTCGTGGACCGGGATGATGAACGACTACAACGCCTCCGCGAGCAACCCGTTA
GTCTGA

>EucaDof16_Eucalyptus_grandis_CDS sequence_reverse
ATGACACAGGCGTTTTTGCTGACACGCGCATTCCACATTCTCTTGTCCTTTGATTTTTCT
CAGGGCACTCTTCACGACGACTCGGGAATGGACACGTCCTCGCCTAACTCAGGCGACATG
CTCGCGTGCTCCAGGCCCTTGATGGAGAGGAGGCTCCGGCCGCAGCACGACCAGCCCTTG
AAGTGCCCCAGGTGCGAATCCACTCACACCAAGTTCTGCTACTACAACAACTACAGCCTT
TCGCAGCCCCGGTACTTCTGCAAGACCTGCCGCCGCTACTGGACCAAAGGCGGCACCCTC
CGCAACATCCCCGTCGGAGGCGGCTGCCGGAAGAACAAGAAAGCCTCTAATTCCGGCAGT
AATAGCAACAACAGTGCCAGCAGCAAGAAATTGGCCAACAATGATCATCAAGGATATGGG
GTCCCTAGCCACCCCGGGCAGTCGTCAGCGTCCTCGTCTCGTGGCGACCACTCGACCGAC
CTCCACCTCTCGTTCTCGCCTGAGGTGCAGCTTTCGCACCTCCACGGCTTGCTCGGCGTC
CATGGGGCCCCCTTCGCGGGCCCGGGCAACAATTTCTCGGGGCTCGAGAACCCGAGGCCC
ATCGATTTCTTGGAGGGCAAGGTAATGAGTTACGACAGCAACTTCATGCTGAGCGGCGAT
AGCCCTCATCCTCACATGGGCGTGGTGGGAGATCTCGGTGGCTCGACTGGGTTGGCCCCG
GGCTTCCATGGCTTCTGCGGTCCATACGGGATGGCCATCGACGGTGGTCATAGCGGAAAT
GCTAACAGTTTCATGGAGAATTGCCAGAGATTGAACATGCTCGTTCCATACGATCAGGGC
CATGATCAGGAGCAGATCAACGCGATGGATGTGAAGCCGAACCCGAAGCTCTTGTCGCTG
GAGTGGCAAGACCAAGGCTGCTCCGACGCGAGCAAGAACGGGTATGGCTATTTCAACGGT
CTAGGACCGTCGTGGACCGGGATGATGAACGACTACAACGCCTCCGCGAGCAACCCGTTA
GTCTGA

>EucaDof16_Eucalyptus_grandis_protein sequence
MTQAFLLTRAFHILLSFDFSQGTLHDDSGMDTSSPNSGDMLACSRPLMERRLRPQHDQPL
KCPRCESTHTKFCYYNNYSLSQPRYFCKTCRRYWTKGGTLRNIPVGGGCRKNKKASNSGS
NSNNSASSKKLANNDHQGYGVPSHPGQSSASSSRGDHSTDLHLSFSPEVQLSHLHGLLGV
HGAPFAGPGNNFSGLENPRPIDFLEGKVMSYDSNFMLSGDSPHPHMGVVGDLGGSTGLAP
GFHGFCGPYGMAIDGGHSGNANSFMENCQRLNMLVPYDQGHDQEQINAMDVKPNPKLLSL
EWQDQGCSDASKNGYGYFNGLGPSWTGMMNDYNASASNPLV*

>EucaDof17_Eucalyptus_grandis_genomic sequence_reverse
ATGGATACTGCCGAGTGGCGACAGGTAAAGACAAAAAACATCTGTTGTTGTTTCTCTTTC
TCTGATCATCATCATCATTTAATTTCCCTATGATTATGTGTCTGTGTCTGCGTGTGTGTG
GGAGATTTTACTCGAGGCGAGCAGAAGAAAAAGTTTCATCTTTGAAAAAGCTTTTCGTTG
ATCTTATCTGCTCAAAAAAGTTAACTTTCTTTCCACCCATGAAATGATGCTTAGCTTCCC
ACTAAATTCTACTACTTTGAATATTGTGTGTGTGGCTCTCTTTCTTAATAAGAAATCTTT
CATCATCATCATTATCTTACTGTTGAAAGTACTGTTCTCCTTTTCTCTCCCTTTCTCTTT
CCCACCTCCTGATAAAGAGATCTGGGAAAGTTGTTCCTTCTTTGTTCTTTTTCCCTTGTA
AACATGTAATCACCCCCCTCAAAGTTGTCAAAAGTTAACTACTTTTGTTTTCTTTTCTCT
GTCCTGTGTGTTTTTATCTTTTCTTCTATTTCTCTCTCTCTTCCCCACAAATCCTGAGTG
ATCCAATCTTTTTCATGCTCTTGGATGATGAAGTCCTAGACTTTTTTCAGTTTTGTTTGG
CACTTTTAACTGTTCTCTTCTTTTCTTTTTCTTTTTTTGCTGAATGATGTTCAGGGAATT
GGGGTTATTAAAGCCATGGAGGGCTCATCGAGGCCTGCCCTAAGGTCGAGGCCTCAAAAG
GACCAAGCTTTAAACTGCCCAAGGTGCACTTCCACACACACCAAATTTTGTTACTACAAC
AATTACAGTCTCTCTCAGCCCAGATACTTCTGCAAGACCTGTAGGAGGTACTGGACTGAG
GGTGGTTCCCTCAGAAATGTACCTGTGGGTGGTGGGTCAAGGAAGAACAAGAGACCCTCA
TCATCTTCGCCTTCAACTTCATCGTCATCATCAAAGAACAAGATTATTCCCAATCTGACT
CCACCTCCTCCAACCATCCTTCACTCTGACCCTCAAAACCCTAATAATAAGCCCCAGGAA
GGCCAAGGTTTTGGCCTGCCCAGCTACTTGCCGGCTCAGGCATACGCTGCCCTTAGCCAC
CACCTTCCAAACCCTAGCTCTTCTTCTTCCCCAATGGAGTTTCTAAAGAGCGGTGGGAGT
TGGCCTAGCGGTTTGAGCCTGAGCAGTTTAATGGCGATTCCCGGGTCTGATTCAAGCGCC
TGCATGTATTCAGCGTCAGCTGGGTTTCCTCTGCAAGAACTAAAGCCAAACCTCCAATAT
TTCTCTCTTGCTGGCCTCGAAGCTGGGGAGTTCAAACAAGTGCCTTCGAGGACTGCTGAA
CTCGAGCAAATGAATAGGGGGCAACATGGAGATCCATCAACTGGGTATTGGAGCAAATTG
ATGGGTGGAGGATCATGGTGATGACATTGTCATATGACAGAAAGATAGCAGCTGTTTTTG
CATGAAAAGCATTTGCAAGAATTCAGGATGAGTTTGATTTCTTTTTGATTTATTTTTTCC
TTTTTAAATATGTCCCTCTCAGCGTGTGGCCTGTTTCTAGGATGATTGAAAGCTAAGATA
TCGAATAGGGGTCAAATGGTTTCTTTATTTTTCTTAATTATTGTCATTCCACAGAGTGAG
TTTGGCTGCTTTGCTTTCCTTCATTTCAATCGTTTCTTATATTCTTCCATGCTCATGGTT
TTTTGTTCACCTAGTCTTGGCAACTTGATCAATGGGGATATGTTTGAGAGGTTTCTTAAT
GGAGAAATTAGTGCATCCTGGAGAAATCCAAAAGGAAGTATATGCATCACGTAA

>EucaDof17_Eucalyptus_grandis_CDS sequence_reverse
ATGGATACTGCCGAGTGGCGACAGGGAATTGGGGTTATTAAAGCCATGGAGGGCTCATCG
AGGCCTGCCCTAAGGTCGAGGCCTCAAAAGGACCAAGCTTTAAACTGCCCAAGGTGCACT
TCCACACACACCAAATTTTGTTACTACAACAATTACAGTCTCTCTCAGCCCAGATACTTC
TGCAAGACCTGTAGGAGGTACTGGACTGAGGGTGGTTCCCTCAGAAATGTACCTGTGGGT
GGTGGGTCAAGGAAGAACAAGAGACCCTCATCATCTTCGCCTTCAACTTCATCGTCATCA
TCAAAGAACAAGATTATTCCCAATCTGACTCCACCTCCTCCAACCATCCTTCACTCTGAC
CCTCAAAACCCTAATAATAAGCCCCAGGAAGGCCAAGGTTTTGGCCTGCCCAGCTACTTG
CCGGCTCAGGCATACGCTGCCCTTAGCCACCACCTTCCAAACCCTAGCTCTTCTTCTTCC
CCAATGGAGTTTCTAAAGAGCGGTGGGAGTTGGCCTAGCGGTTTGAGCCTGAGCAGTTTA
ATGGCGATTCCCGGGTCTGATTCAAGCGCCTGCATGTATTCAGCGTCAGCTGGGTTTCCT
CTGCAAGAACTAAAGCCAAACCTCCAATATTTCTCTCTTGCTGGCCTCGAAGCTGGGGAG
TTCAAACAAGTGCCTTCGAGGACTGCTGAACTCGAGCAAATGAATAGGGGGCAACATGGA
GATCCATCAACTGGTCTTGGCAACTTGATCAATGGGGATATGTTTGAGAGGTTTCTTAAT
GGAGAAATTAGTGCATCCTGGAGAAATCCAAAAGGAAGTATATGCATCACGTAA

>EucaDof17_Eucalyptus_grandis_proteomic sequence
MDTAEWRQIGVIKAMEGSSRPALRSRPQKDQALNCPRCTSTHTKFCYYNNYSLSQPRYFC
KTCRRYWTEGGSLRNVPVGGGSRKNKRPSSSSPSTSSSSSKNKIIPNLTPPPPTILHSDP
QNPNNKPQEGQGFGLPSYLPAQAYAALSHHLPNPSSSSSPMEFLKSGGSWPSGLSLSSLM
AIPGSDSSACMYSASAGFPLQELKPNLQYFSLAGLEAGEFKQVPSRTAELEQMNRGQHGD
PSTLGNLINGDMFERFLNGEISASWRNPKGSICIT#

>EucaDof18_Eucalyptus_grandis_genomic sequence_forward
ATGATTCAAGAGCTGTTGGGAAGTCCTGCACTTATAGGAGAGAGGAAAATCTCCATGAAC
CATGGCGCGGTTATTTTAGAGGGCATGCCTTCTCCTTCTCCTTGCTTCTCTCCATCTCCC
ACGCCTTCGCCTCCTTCTTCTTCCTCCTCTGCCATCACAAAGACGACAGCACAAGCGACG
GCAACGGCAACAACCACCGGTAATGGGACTTCGACAACAACGAACTCATCAAATCCCGAG

AACCTGAGGTGCCCACGCTGTGATTCTGCAAATACCAAGTTTTGCTACTACAACAACTAC
AACCTCACGCAGCCTCGCCACTTCTGTAAGACGTGCCGCCGCTATTGGACCAAAGGAGGA
GCACTAAGGAATGTTCCTATCGGAGGAGGATGCAGAAAGAACAAGAGCTCAGCTGTGTCA
TCCTCTGCCTCCAAATCGAGCGCTTCCAAGTTCAAGAACTTATCGTCTGAGATCGGACGG
TCGAGTTTTGTCAGTGCTTTCGATCCTGAAATCTCACAGAGCCAGGTCCTTTGGGCCTCA
CCCCAGAGTTCCCACATCCTGGCCCTCCTGAGAGCCAATCAGAACCCTAACCCTAGTAGC
CTTGGTCACCAAGCCTCTCCCTTTATGACTCTGAAGGAAGAGACCAATCTGATTGGAAAC
CACATGATGCCTGAGCCAGGAATTTCAGAGAGTGCAATGAATAATGCGAGAAACCTGGGC
ATGATTGACCCTCTCATTCATGTTCCTTCCTCTTGCACCTCCTTTTTGAGAGGTAACCAT
CAACAAATCCCACACCAACTCAACCATCACCAACAAAATGGTATTATCCTCAGTGGTGAA
GCTCAAAACTCCGGGATACAATCGCTATACCAAAGGCTCAGAGCTTCCACAGGTAATAAT
TATGGTTACAGTGAGCACTCGCCGGCAATTCTAAGCAGCATGGCAAGTACCCCGCCGCTG
TCATCTTCCCTGTCCTCGTCCATTCTAGAGTCAGTCCCAGCCACGGCGACAAGTGAACTC
GGGTACTGGAACCCGGCTTTCTGTGTGTCGGATCTTCCCAACACAAATGGTTCATATCCT
TAG

>EucaDof18_Eucalyptus_grandis_CDS sequence_forward
ATGATTCAAGAGCTGTTGGGAAGTCCTGCACTTATAGGAGAGAGGAAAATCTCCATGAAC
CATGGCGCGGTTATTTTAGAGGGCATGCCTTCTCCTTCTCCTTGCTTCTCTCCATCTCCC
ACGCCTTCGCCTCCTTCTTCTTCCTCCTCTGCCATCACAAAGACGACAGCACAAGCGACG
GCAACGGCAACAACCACCGGTAATGGGACTTCGACAACAACGAACTCATCAAATCCCGAG
AACCTGAGGTGCCCACGCTGTGATTCTGCAAATACCAAGTTTTGCTACTACAACAACTAC
AACCTCACGCAGCCTCGCCACTTCTGTAAGACGTGCCGCCGCTATTGGACCAAAGGAGGA
GCACTAAGGAATGTTCCTATCGGAGGAGGATGCAGAAAGAACAAGAGCTCAGCTGTGTCA
TCCTCTGCCTCCAAATCGAGCGCTTCCAAGTTCAAGAACTTATCGTCTGAGATCGGACGG
TCGAGTTTTGTCAGTGCTTTCGATCCTGAAATCTCACAGAGCCAGGTCCTTTGGGCCTCA
CCCCAGAGTTCCCACATCCTGGCCCTCCTGAGAGCCAATCAGAACCCTAACCCTAGTAGC
CTTGGTCACCAAGCCTCTCCCTTTATGACTCTGAAGGAAGAGACCAATCTGATTGGAAAC
CACATGATGCCTGAGCCAGGAATTTCAGAGAGTGCAATGAATAATGCGAGAAACCTGGGC
ATGATTGACCCTCTCATTCATGTTCCTTCCTCTTGCACCTCCTTTTTGAGAGGTAACCAT
CAACAAATCCCACACCAACTCAACCATCACCAACAAAATGGTATTATCCTCAGTGGTGAA
GCTCAAAACTCCGGGATACAATCGCTATACCAAAGGCTCAGAGCTTCCACAGGTAATAAT
TATGGTTACAGTGAGCACTCGCCGGCAATTCTAAGCAGCATGGCAAGTACCCCGCCGCTG
TCATCTTCCCTGTCCTCGTCCATTCTAGAGTCAGTCCCAGCCACGGCGACAAGTGAACTC
GGGTACTGGAACCCGGCTTTCTGTGTGTCGGATCTTCCCAACACAAATGGTTCATATCCT
TAG

>EucaDof18_Eucalyptus_grandis_protein sequence
MIQELLGSPALIGERKISMNHGAVILEGMPSPSPCFSPSPTPSPPSSSSSAITKTTAQAT
ATATTTGNGTSTTTNSSNPENLRCPRCDSANTKFCYYNNYNLTQPRHFCKTCRRYWTKGG
ALRNVPIGGGCRKNKSSAVSSSASKSSASKFKNLSSEIGRSSFVSAFDPEISQSQVLWAS
PQSSHILALLRANQNPNPSSLGHQASPFMTLKEETNLIGNHMMPEPGISESAMNNARNLG
MIDPLIHVPSSCTSFLRGNHQQIPHQLNHHQQNGIILSGEAQNSGIQSLYQRLRASTGNN
YGYSEHSPAILSSMASTPPLSSSLSSSILESVPATATSELGYWNPAFCVSDLPNTNGSYP
+

>EucaDof19_Eucalyptus_grandis_genomic sequence_forward
ATGGTTTTGGCATTGAGGGCGCGCGAGAGAGATTCTGCGGATAAGATGAGCGGCGGAGGC
TTCGTGTCCGAGAATAAAGATCCGGGCATCAAGCTCTTCGGGAGGAATATTCCGGTGCCG
GACTGCCGGGTTCCGGCGGGCTCCGCTGGCGGTTCCCGGGCTCTCTCGGACTCCGAAGTG
ACGGTACAATTTTTCTCTTTTGCCCCAGCTCTAGGAAAGCTCGTGTCAGTGCCTGTGGAT
TGTTTCGCGTTCTTGGATTGTTTGGAGCTTTAGTTGAACGTCTATGGTGACCGACAACGT
GCTGTTTTGCCTCTTCTTGCCTTTAAAATGGAATGAAGATGTACAAGGAAATATTGAATC
GCTTGCTCAAGAGTTCAACGTTGTTGGGTTCTCCGTTCTTTAGTCTCTTTCTTGTAAATT
AGCCTTTTCATTGTGTGATTATGCTGAATTGGAAGATTGTTCTGGTTGTGTATTGCTGTC
AAATGACAATTAATCAGTACTCTGCTGCATCTATTTTGTGGGCCTTGTACTAACACAGCA
CCAACATGCAATAACAATTGGCTCGGTCTCTGTTCTTTGGTTTTCTGCACAATAGAATAT
GATATTTTCTTTCTTTTTTATACGTGGTGCCATTTGCTAACTGTGCTCACCCATTAAAAC
CAATTCTTCCAAGCGCCGAGATTATTCATACTTGGCTAAAGAGCATAGTTGCCTAGAGCA
TCGACAGTCTTTTAAGTGAACATTTTAGTAAATATATGTAGATGATTAGTCCTGGTTCCA

CCAAAAGAGCATTGATTTACCAGACATCTGAAACATGGGATTTTGAGTGACCCGTGAACT
TAATTGAAAATTGTTCTGCTCAGTCTTTGAATTTTTGTGCTATCAACAAACTTGTTTTCC
TGTTATTCATTCTGCCATAGTGGAATTTCCTGCTTCAATTGAGTTCAAGATTTGATTATA
ATTTGTTCTAAGATGTCAGAAAATATTACTGTGCATTAAGAAGAAAGATCATGACAGACG
AAGAATTGGATGCTAATATTGTTATTGAATTGTTAGGTTTGCTGTTATCTTGTTAAAATA
TTTTCTTTCATTTCTGGTTATTTCTCTCTTTGATGTACTTAGTTTGGACGAGATCAATAC
TTTATGTTGTTCTAATTTAGTAATTTTCAGCATGCCTGTAGTGATATTGAAAAACCAGAA
ACTGTCAAGTATGGACATCCTCATAATTCAGCTGTTGGTGGAGAGAAGGCCATGGTAATG
CGGAATCTTCAAAATGGTCATGCTGAGACTAAGTATGCTGACCAAGAAAAGGTTTTCAAG
AAGCCTGACAAGATCCTTCCATGTCCCAGATGCAACAGTCTTGAGACGAAGTTCTGTTAT
TTCAACAACTACAATGTCAATCAACCTCGACACTTCTGCAGGAACTGCCAGAGATATTGG
ACAGCTGGGGGAACAATGCGAAACGTTCCAATTGGTGCTGGACGGCGGAAGAACAAGCAC
CTGGCTTCTCAGTATTGTCAAGCGACTCTATCTTCCAGTAGCACACTAAACACTAGATTA
GACATTCCTGAGATGACAGTGAATCATCAGTTTATGGCACCGAAGGAAATTACTTCAACA
ACACTTGGGCCTTCAGTGGGAAATGCTAGAGTATTGAAATTTGGTCCTGAGGCATCTTTA
TCCGAATCTGTTGAGACTGTTCTTAGTCTCAAAGACCCAAAAATATCCATTGAGACTGGT
TCAATCAATTGTGTCCAAACTAGAGAAGAAGTTTCCTCATCTGGTTCTTCCATGACAACC
TACAGTATTCATGATTCGACCAAAAGCGCTGCTGAAAATGAAGAAATAGGGATATCGGGA
TCTCGTCGCGAGCAGAATTCCCTCGATCCTATGCATTTTTACCCTGTTCCTGCATGTTTT
ATCCCCTTGAACACAGGTTGGCAACATCTAGCTTCTTTTAGGACAACTCAGTACTCTCCA
GAGCAAGTTGCAGGTGTTGCTAGCCCCAATTCCGTTCATTTGAGCTCCACACCCATGTTA
GCGGTTCCTGGTATTTGCCCTCCTAGCATTCCCTTACAAATTGTACCCATGGCGCCGTAT
TTTGGTTGCATGCCTCTACGCCCCTCTGGGCTGAGCACCACACCATCAAGTGGAATGAAC
AGTGGCCCATCATCTTCCTCTCCATCTTCCCCAATGGGTAACGCATATTGCTCTGCTAGT
GCCTCGCCAACCCTTGGCAAGCACTCAAGAGATGCAAATTTTGCAGATTTTGAAAGATCA
GATAAGTGCATCTTGGTTCCAAAGACACTGAGGATGGATAACCCAAGCGAAGCCTCGAAG
ATTCCTATATGGGCGACACTGGGTATTAAGCCAGGTAAGGCTAAAGTGGAAAATGAACTC
AATGATTCCGATGCCACACACATGATGGAAGTGAACCCAGCTGCATTTTCTCGGTCTCAA
ACATTTCAAGAGAGCACATAAGAGATTGTGATGAATTTGTTGTGATTCAGTCAGCTTGTA
AACAAATACTGATAATGGTGAGCTCTGTTCCAAAAGCTGCTGGGTATGCTCACTGCTTAC
CATTTGCGAACTTGATCTGTATATATTCACGTGAGCCAGAATGAATGAAAGATTTCAATA
ACATCCAAAGCTTCTGTAGTAAGTGTCAATCAGGGAAAAATGCACTTTGTCTACTTGTGT
AGCATGTGTGCTCGGTCTTGAAACTTGAAAGCAGTGGAATTTTTTCCTCTCCTGGAGATT
TGACTGTCTCTCTGCATCAGCAATGTTGCAGCTTTACATAAAGCTAGTTAATCATTTACT
CAAGAGGCATGTTAAGGCATTGGCTACTGGTTTATTCCCTTGGATAATGATCGGTTTTGC
TTGCTTAATCCCGATCTCAATGCTGTCAGGTTCTCTTTCTTGTGCTTTATCTGCAGATCAA
CTACTTAACATTCTCGACCACTTATCATGATGGGCCTCTACCAGATTGGCACTGTACTTT
CGCGTCTAGGGCGGCCTTCATTATTGGCCGGGGTTGCTTACTGCAATTCCTTTAGATATG
ATCCGGGTATCAGTAAGACCGGATTCTTCAGAGCATCACATCTTCGTGCTGCCAACAGTT
ATATGTATCACGTAGGCTTATCCATTTGTCGCACTTGAGCGGAACGAAGTTTCGAAAAGA
ACATGATGCAGACTCAGCTCCTGCTCCAATTCCCGTCTAGTTTGAAGCTCTGGACACAAC
ATTGTGAATCAGAGTGCGATGCGACTTAGTTTATTTTGGGTTCTCTTTCTTTCGCTTTGG
TTCCAAAATCGAAACCGGGTGCGACAACACGCGGGTTCTTCCCTGCTTCCTATGCTCAAT
TACTTGAACACTCCAACTCCCGCGGGATTTCAGAATGTTTCGCTGGATGGATTCTTCGGT
GCATGAGGCACCGCAGAACAGAGTGATGACTGACCAGTGGCTCGGCATTTCTTCTACATA
AAATTTGCAGACATCGTCATTTCTCGAGTTAAAGACATCGAAGCGGACGTTAACAATGTT
CTGCCACTTCAAAATTTCAAATTGCGAGGGGCAAAGATGGGCTTTCAGAGCCCGTTCGTG
TACGACGCATTGCACGACAGAGACCACTCGCAACACTTGCACGACAAAATCTTACATTAT
CTTGTGATACTTCCTGAAACTGGAGCGTTTCATTATTTTGCGAAGGGGAGACATAGAGAG
CTCTCTCCTGCAAGCGCTAGACAGAGAAACACCCAATTAACCAAACTTTGCCGGTTCAAA
ACCTTAAGCTTCAATGATCTTCACCGCCTCGAAATCGAGGCTTCGTCGGGTCAATTTCCG
AGCCTCAGCAAATCCTCATCCACTCATGCGGACAGCTCCATTGATCTCCATCGACTCAAG
TCTCGCAACCCTCATAAGGTTCGATCACTGAGACCGGGATTCGCCCGGATCAAGTCATTA
AGGCCTCGAGCATCTTAA

>EucaDof19_Eucalyptus_grandis_CDS sequence_forward
ATGGTTTTGGCATTGAGGGCGCGCGAGAGAGATTCTGCGGATAAGATGAGCGGCGGAGGC
TTCGTGTCCGAGAATAAAGATCCGGGCATCAAGCTCTTCGGGAGGAATATTCCGGTGCCG
GACTGCCGGGTTCCGGCGGGCTCCGCTGGCGGTTCCCGGGCTCTCTCGGACTCCGAAGTG
ACGGTACAATTTTTCTCTTTTGCCCCAGCTCTAGGAAAGCTCGTGTCAGTGCCTGTGGAT
TGTTTCGCTAATTTTCAGCATGCCTGTAGTGATATTGAAAAACCAGAAACTGTCAAGTAT

GGACATCCTCATAATTCAGCTGTTGGTGGAGAGAAGGCCATGGTAATGCGGAATCTTCAA
AATGGTCATGCTGAGACTAAGTATGCTGACCAAGAAAAGGTTTTCAAGAAGCCTGACAAG
ATCCTTCCATGTCCCAGATGCAACAGTCTTGAGACGAAGTTCTGTTATTTCAACAACTAC
AATGTCAATCAACCTCGACACTTCTGCAGGAACTGCCAGAGATATTGGACAGCTGGGGGA
ACAATGCGAAACGTTCCAATTGGTGCTGGACGGCGGAAGAACAAGCACCTGGCTTCTCAG
TATTGTCAAGCGACTCTATCTTCCAGTAGCACACTAAACACTAGATTAGACATTCCTGAG
ATGACAGTGAATCATCAGTTTATGGCACCGAAGGAAATTACTTCAACAACACTTGGGCCT
TCAGTGGGAAATGCTAGAGTATTGAAATTTGGTCCTGAGGCATCTTTATCCGAATCTGTT
GAGACTGTTCTTAGTCTCAAAGACCCAAAAATATCCATTGAGACTGGTTCAATCAATTGT
GTCCAAACTAGAGAAGAAGTTTCCTCATCTGGTTCTTCCATGACAACCTACAGTATTCAT
GATTCGACCAAAAGCGCTGCTGAAAATGAAGAAATAGGGATATCGGGATCTCGTCGCGAG
CAGAATTCCCTCGATCCTATGCATTTTTACCCTGTTCCTGCATGTTTTATCCCCTTGAAC
ACAGGTTGGCAACATCTAGCTTCTTTTAGGACAACTCAGTACTCTCCAGAGCAAGTTGCA
GGTGTTGCTAGCCCCAATTCCGTTCATTTGAGCTCCACACCCATGTTAGCGGTTCCTGGT
ATTTGCCCTCCTAGCATTCCCTTACAAATTGTACCCATGGCGCCGTATTTTGGTTGCATG
CCTCTACGCCCCTCTGGGCTGAGCACCACACCATCAAGTGGAATGAACAGTGGCCCATCA
TCTTCCTCTCCATCTTCCCCAATGGGTAACGCATATTGCTCTGCTAGTGCCTCGCCAACC
CTTGGCAAGCACTCAAGAGATGCAAATTTTGCAGATTTTGAAAGATCAGATAAGTGCATC
TTGGTTCCAAAGACACTGAGGATGGATAACCCAAGCGAAGCCTCGAAGATTCCTATATGG
GCGACACTGGGTATTAAGCCAGGTAAGGCTAAAGTGGAAAATGAACTCAATGATTCCGAT
GCCACACACATGATGGAATCAGCTTGTAAACAAATACTGATAATGGTGAGCTCTGTTCCA
AAAGCTGCTGGGTATGCTCACTGCTTACCATTTGCGAACTTGATCTCAATGTTGCAGCTT
TACATAAAGCTAGTTAATCATTTACTCAAGAGGCATGTTAAGGCATTGGCTACTGGTTTA
TTCCCTTGGATAATGATCGGTTTTGCTTGCTTAATCCCGATCTCAATGCTGTCAGGGCGG
CCTTCATTATTGGCCGGGGTTGCTTACTGCAATTCCTTTAGATATGATCCGGGTATCAGT
AAGACCGGATTCTTCAGAGCATCACATCTTCGTGCTGCCAACAACATCGTCATTTCTCGA
GTTAAAGACATCGAAGCGGACGTTAACAATGTTCTGCCACTTCAAAATTTCAAATTGCGA
GGGGCAAAGATGGGCTTTCAGAGCCCGTTCGTGTACGACGCATTGCACGACAGAGACCAC
TCGCAACACTTGCACGACAAAATCTTACATTATCTTGTGATACTTCCTGAAACTGGAGCG
TTTCATTATTTTGCGAAGGGGAGACATAGAGAGCTCTCTCCTGCAAGCGCTAGACAGAGA
AACACCCAATTAACCAAACTTTGCCGGTTCAAAACCTTAAGCTTCAATGATCTTCACCGC
CTCGAAATCGAGGCTTCGTCGGGTCAATTTCCGAGCCTCAGCAAATCCTCATCCACTCAT
GCGGACAGCTCCATTGATCTCCATCGACTCAAGTCTCGCAACCCTCATAAGGTTCGATCA
CTGAGACCGGGATTCGCCCGGATCAAGTCATTAAGGCCTCGAGCATCTTAA

>EucaDof19_Eucalyptus_grandis_protein sequence
MVLALRARERDSADKMSGGGFVSENKDPGIKLFGRNIPVPDCRVPAGSAGGSRALSDSEV
TVQFFSFAPALGKLVSVPVDCFNFQHACSDIEKPETVKYGHPHNSAVGGEKAMVMRNLQN
GHAETKYADQEKVFKKPDKILPCPRCNSLETKFCYFNNYNVNQPRHFCRNCQRYWTAGGT
MRNVPIGAGRRKNKHLASQYCQATLSSSSTLNTRLDIPEMTVNHQFMAPKEITSTTLGPS
VGNARVLKFGPEASLSESVETVLSLKDPKISIETGSINCVQTREEVSSSGSSMTTYSIHD
STKSAAENEEIGISGSRREQNSLDPMHFYPVPACFIPLNTGWQHLASFRTTQYSPEQVAG
VASPNSVHLSSTPMLAVPGICPPSIPLQIVPMAPYFGCMPLRPSGLSTTPSSGMNSGPSS
SSPSSPMGNAYCSASASPTLGKHSRDANFADFERSDKCILVPKTLRMDNPSEASKIPIWA
TLGIKPGKAKVENELNDSDATHMMESACKQILIMVSSVPKAAGYAHCLPFANLIMLQLYI
KLVNHLLKRHVKALATGLFPWIMIGFACLIPISMLSGRPSLLAGVAYCNSFRYDPGISKT
GFFRASHLRAANLPLQNFKLRGAKMGFQSPFVYDALHDRDHSQHLHDKILHYLVILPETG
AFHYFAKGRHRELSPASARQRNTQLTKLCRFKTLSFNDLHRLEIEASSGQFPSLSKSSST
HADSSIDLHRLKSRNPHKVRSLRPGFARIKSLRPRAS#

>EucaDof20_Eucalyptus_grandis_genomic sequence_forward
ATGTCCGGTTCGCTTGAAAGCATGCTGGTCTGCACGAAACCCCAGCAGGACAACAAGAAG
CCCAGGCCTCAGCCAGAGCAAGCCCTAAAGTGCCCGAGATGCGACTCCACCAACACCAAG
TTCTGCTACTACAACAACTACAGCCTCTCTCAGCCTCGCTACTTCTGCAAGTCCTGCCGG
AGGTACTGGACTAAAGCGGGACCCTGGGGAAGGTGAGAACAGGATCTTGTGGGGATTCCC
ATGGCAGCTCAATGCCAATTATGCAGATGGAGGGAACATGATGGGTGAACTTGATTCTGC
AAGAGGAAGCGGCTGGAGCAATGGGCTCAGCTCCCCCATGGCATGGCCTTTCTTAATAGC
CCTCTGATGTAGGAACTCAACAAGATTTTCCAAGTTGAAGTGTATGCAATTAGTTTAAGA
AGTGGCCAGAGGGCATTAGGGGGTTCCTCCTGTTCCAGAATGAATGAATCCATTGTCTTA
AAGGAATTTGGTCCTCTGTGTGGTGTGCTAAGGGATGAGATTAGTCTAAAATCATCTCCA
ATTTTTCTTGGGGGTTTTGCCTTTTCTTTTAAAATCTTTAATCTTTCCTTGTTTAATTTA

ATGAGAGAATGGAATTTCGATTTTCCTTTTTCTAAGAGATTGTATTTGTACGGTGTGACT
CCAATGGTGGTGGCCAATCAAGTGGTCCAAAACTCAATTTGCAAGTCCAAAGTAACAAGT
TCTGTTCTCTAGCCTACCCTTTAACTTGGAAAAATCTTTATTAGTCATTGAAAGACGAGT
TTGTACTTAGGTGTACTCTCTATTTTGAGTAACCCACTAACATCATAATGGGGTAAGGAT
GATAAACTTGCATAATGAGCATTGTACGTATATAAAATCCCCAATCTCAAATCGAAAATT
CTACTAGGGGTGATTCATACCTTGTCAAATGCTAGCGCATCCTTTTATGGGTTTAAAAAT
AGTATGTGAATGCATGAGCCAATTATATGTGACTGTTGTTTACGTTTCATGATCTAGGGA
ACGAAATTTCAAATTTGTCAATCTCAAACTAAAAATTCTACTAGAGGTGATTCATACCTT
ATTAAATGCTAGGCATCCTTTTATGGGTTCAACAATAACATGTGAATGCATGAGCCAATT
ACACATGATTTTTATTCACGTTTCATGATCCAAGTAATGAAATATCGAATTAGTCGTTAC
AGCGAAATCACAAGCAAAAACAGTAGATGTTGTTATGGAGAAGTAATTCAAAATTAAAGC
TTGAAAAAGTTCTCACGAAATCTATTATCTAGAGATTGGCCGTCGATTCTCTATCCTTCC
GAGCTCTTTCCGGGTTTTGGAGATAAGAGAAAAAGCAGGGACAACGAGTGACGTTACAGG
CGTTTTGTGCAGCCTGGGACTTTTGCTTTACGTACTCTTATCCCCCTGCAGGTCGTTCCA
GAAGCGCATTAATGTTTTTCCCTTATAATGCGAGCGGGCAGTTCTCGAGATATTCCGTCA
TGATTGCCCTGGCTGTTCAGACAAGATTTCGAGAAGATCAAGAATTCGGATTATAATAAG
GAAATAATTGTCCGATGATGGGGTCGATACAAGTCGCTCCTTTTCCGCCCGTCCTATTTC
TTTTCTTTTAATATATCACGTTTTCTCCCCGCAAAGCATTTCGGGATTTCAGTGAAGCGT
GAGCGCACCGATTCCGGAGAGGATTTGCAGTTCCTCCGCTTCCCTTCTTGCTCTTCTGCC
CAGACAAGATCCAGCAACAAAGATTCGGGCGTGCCTGAGCGGGCAGCGCGATCTGGCGAG
GACTGCTGACTCGAGAACGAGAGGAGGACAGAGTACGAGATTCCAAAGCGGGGGGAGGCA
GCCCGAGATGTCAGAATCGGAGAGTGAAAAATAATGCGCAGGACATCTGCTCTGCTTGAC
TTCAGTGTCTGTCTCCGCTTCTTTTTTTGACTATCTGTCGTCCTCCTGGGTCCGGTCTCG
TCTGGATGTGAATTGGGTTGGCTGGGATCGAGGAATGAGTTGCTTTCGATGGACTAATTG
TCTGAAGCTGGATGAGCTTTAAAAAGCTCAAAGCTTGAATGAAATAGTACTTAGCTAAAA
AATCAAACAGCTAAGCTAGATAGATCAATGTTGTGTGTAAACAGACGTACGTTGAGCTTT
AGTACCAAGTGATTAATCCGTGACATGGGTACCGTAAAGAATGCCTCAGGTTGACCTATT
GGTGATAAACATTGGCGGACTCGTGATCGATTTGTATCTATATCTGTTTCCATATCTATC
TACACGGTTATCTAATTTTGCATAAATTCATGTTGGGCTCCACAATGATGATTGATATTT
GGACTTCTTATGTGAGGGTCAGGTCTTTCTCCTTTCGAGGAGTACTTTCCTTGTGTTCTC
CACTCATTTTGCTTCTAAGTGAAATGTCTGCATGCTCCGTTCAACAAAAGAAACCATAAG
TTTTGATTTGTTTTAGTGTTGCAAATTAGCAATTATGAAAATTCTAGCATTTCGCATGGC
ATTTTTATATGATAACGGTATATGCAGCCCATTTGTTACGCAATCCATCTTCATCACCAT
TGATATAGGTAGTTGTAGTTTGATACTCTGACAAATTATCTATACAAACGGTAATGGATT
TAGGTTGTGTACAAAATTTTTTATCCGGGTTATGTATTTTAAAATTAATCCTTTCTTGTA
GCTCTTCCTACTGAGTATTCCATTATTGCCCTTCCATAATCAATACCATGTCCTTTCCAT
TGGTGAGGGATCACATATAGAATTTCCCTAAAATAACACAGGCCGATTTACCAATAGAAT
TAGTCTAAAAGGATATAATTTCTTCTCATGTGTTTTGTAATAAAAAAATTAAATATATTT
CATATTTGGATCTTCCAATTTTTTCACGTTTAATCTAGATTTAAAAGATCCATAAGTTAT
ACAAGTTGCTTGATTTCCCTATGACGTCGTATTCGGAAGCTAATCTTAGAAACATTTTAT
ATTTAGGTGGTGTTAGGGGTGAGTATAGTTCAAGGGTATAATGTGGAACTTGAGAACCGA
ACGTATGGGAATTTCTAGGTTCTAGGTTCTAAAGTATACAGGGTAATATCCGGGTTCCAA
AAAATAACGAACTTATTTCAACGGATAAGTTTCAGGTTCTTAGTTGGAATAATTTGTAAC
ATGTAATCTGTAACCTAGGACTTGAACAAGTTTTAATGTTTTTTTGGGTTCATGTCTTCG
CACAATCTTATGATATTTCATGTCATTCATTAATGAGTGTATACTTTAAAAGCACTAATC
AAAACTTCCATTTTATGACTGAGATTTTCACTTTATAAATGTTCATATTTTTCATATATC
TTTATGGTTAGTGAATTGTAGCAGCAAATTGTGGATATTAAAGGTTTTAATTTACTGTAA
TCGTTCAATTAATCATATCAATGGAACTTAGATAGATCTTGAAATCGACTTTGAAACTTG
TGACAGAGTAGCTTCCATGTTTTAGGATATGCAAAGTAGGTTTTAGGTTCTAAAAATGAG
GAACCTATTCTAATGGGTAGGTTCCGGGTTCTAAATGGAACCTGTATAGAACTTGGAACC
GCTCACCCTTAGGTGGTCCCATGGTTCTTCCTATTAGCTAATTCCTTTCATTCCCTTTCC
TTTCACGTCTTTTTTTTATTAGGTTTTATTAATTGATATTAATATATGAAATCTTGAACT
TTTATGCGACGTGTCAACGTTTTTCACTTGATTGCCCCCATATTGCCTTTTTGTTGCCCC
TAAAATGTTAGGAAGGCGATAATCTTTTAAAATAAGACTTTTGTTTTGTACATTGCTGGT
GTAAGAAAATCATCATACAAGCATAGTTCAGGCACCAACACATGTACTAGAAATATATAT
ATTTACAGACACATGTACAATATTAAATTTTGATATTCCACAGCAAATTCATTTGTTCTT
TTTTCTTGATTCTAGTTACATTGATGTTGCCACTGTTAGCATTGTATCTAAAGAGGTTTC
TAAGACTATAGTAATTTTATTTAGTAGCAATTTGAAATGATTTTCAATTGAAAAATATAA
TTCTCCATAATCCCGTTGTGGACTTCATGTTGATTGGAACCGAACATCAATCAACGTACT
CGACGGGTTCCCATCAAACACGCCTTGGTCGTCTACAAAAACCCACAGGTAAGTTCAATT
TTTATGGCGACGTACTTCAACTTTTTCACAAGGAAATATATATGAAAATTAAATTGAAAA

GTGACAACCGACCACAAAGGCGATAATCCGCACATAAGTGAGGTATGTGCTGATATTTAA
ATAATATGAAAATTAAATTAGTAAATAAATAATCACATAAAAAGTTAAAAGTTTACTTAA
ATAAAAGATAACACATCCACAAGCGCGGTAATCCGCACTCAAGGAAAGTATATCTCGTGA
TCTAGGAACCAAGCGATACCTCTTTGGATGAGAAGAGACGGTTATTTATAGGGTGATGAA
TTGTAAAAGGCGTTAAAATGAAGGCACTATAGTGAAAGTGTTACAATAAGAAGGGTGGCT
AGGGACACAATCTCTTTGATCAATGCAAGGTAACACCATTTTTCTAATTCTATTTAATTC
ATATGATCTAAAATTAGTATGCAGATTCTTTTCAACTAAAACATACTTGAAATTCACAAT
TAAACTATTTTCGGATATTATCCCACTTATCATGATTTCTTTTTCACCATACCTTTCATG
ATATCAACGACTAAATACCCGTTCATGTCCCATCGTTTGATTTTGTTAATTAAAGGAATT
GCAGGTGGTGATAATCGATGGAGTTTGCCAATATATTTATGAAATAAACTGAAAAGGAAT
TTCTTCCAATTTTTTAACTTTGGGGATTACATTCATGAAATTTTCACAACAGTAAGGCCA
GTAGCTCTGAACTGAAAGTCACCTCCTTCCATATGGGGTCCGTTTCCCGGGAAGAGAACA
CGTTCAAAAACCAAAGCAAACTTCTTCTTTATGCGACAACATACGACGACATCAAAGAAG
ATCACGACATAACACTTGATACTTAA

>EucaDof20_Eucalyptus_grandis_CDS sequence_forward
ATGTCCGGTTCGCTTGAAAGCATGCTGGTCTGCACGAAACCCCAGCAGGACAACAAGAAG
CCCAGGCCTCAGCCAGAGCAAGCCCTAAAGTGCCCGAGATGCGACTCCACCAACACCAAG
TTCTGCTACTACAACAACTACAGCCTCTCTCAGCCTCGCTACTTCTGCAAGTCCTGCCGG
AGGTACTGGACTAAAGCGGGACCCTGGGGAAGTAAGGCCAGTAGCTCTGAACTGAAAGTC
ACCTCCTTCCATATGGGGTCCGTTTCCCGGGAAGAGAACACGTTCAAAAACCAAAGCAAA
CTTCTTCTTTATGCGACAACATACGACGACATCAAAGAAGATCACGACATAACACTTGAT
ACTTAA

>EucaDof20_Eucalyptus_grandis_CDS protein sequence
MSGSLESMLVCTKPQQDNKKPRPQPEQALKCPRCDSTNTKFCYYNNYSLSQPRYFCKS
CRRYWTKAGPWGSKASSSELKVTSFHMGSVSREENTFKNQSKLLLYATTYDDIKEDHD
ITLDT#

>EucaDof21_Eucalyptus_grandis_genomic sequence_forward
ATGTTCGCACCCTCGAGCGACCAAGCGGCGCTGTTCCAGCGGTGGGGGACCTCAAGCCAC
GGCGGCGTCCCTGCCGCGGCGCCAAGCTGCCCCCGCTGCGCCTCCAGCCACACCAAGTTC
TGTTACTACAATAACTACAGCCTCTCTCAGCCCCGATACTTCTGCAAGTCCTGCCGCCGG
TACTGGACCAAGGGCGGCTCGCTCCGCAGGGTTTTACACTGATTTGAGGTGTACATCAAT
ATGTTGATTGATGTTTCGCTTTAGCTTAGTAGAGAATGGGATTTTTCTATCTCTAGCATT
TAGATTCTTTTTTTTCTTTGCAATTTATTTAATGCTTTCTTAATCAATGATTGTTATTTT
AATAATTGTGCGCATGCATGATCACCTCATATATCAGTGGATAAACTTTAAATCAATCCA
CAATGTTTGACAAACATCTTCTTGAATTGAATGAGATTAGATATCGAAACGGCACTAATT
AATTAATTAAAGTAACCAAGTAGTCAATCCTAAATTTTCTAGTTGTACAAGAAATGATGT
ATACTCGCATGCCTCACTTGATTTCTAGATAATCTCAATAGCCTAATGGTGGCTACTCGT
TAGAAAATTCTTTTACCATGAATTGAAGCTAAAATTAAACACCACAACGTTACGTGAGGT
TTGGGCTTGGGAGAGCCTACACTTAATTGAGGATTATAGACCCATCTTTTAGGCAATATT
ATAAAACCTATGTCTTTTCCCCTCTGAACAAAAAAGGCAAGTTGTGACAACTATCACAAA
GCAAATCTGATTTATATTTCGATGTAATACTTTTGAAGAATGATTTGCTTCATTCTTTTA
AATCTACATGGTTATGTTTACTCCCTTATGCAAAGTCTTTCCATTTCATATTATCTGTGA
GGATCCTCTGCTTAAAGAACAATTCATTCAAAAAAGCCATATGTACCATTTAGGAAATCA
TTAAACTTCACTTATTTCAGTCCCTTAATGAGTCTTTTCATGTTAAATCATTATCATGTT
TTTGATACACTGCAATATTTCTCACATGATTCTCTATAGCCAATATGTCATACTATAGAG
AAATGTTGCTAGCATATGACAGGAAAAATTATCAACTATTTATCTTGTCCTACTATGGCA
AAAGTTGGGTAGTCGGGGAGCAAATAATATTGATGCAACGAGGTGAGGGGATCGTTATTT
GGGATTCCTAAGCACGATAAGTAGTGCGTGATACAACATGGCTAATTACATGATTAATCG
AATTGCATAGTTGATTAAATGATCGTTGCACCTAAACATATGTGATTTAACAATCAATTG
GGTAGGCTCATCGTCCGTATTGTTTTGCTTTCCTTTCTAGGGCAACCAAAGATAAATGCA
CGAGTGGCGTACCCCAAATCTGAGAAATTTCAGTTAACCAACATTTTCCTTTTTTTTGAA
AATTAGTAAAACCAAGTTATCATATAGTCGTTCCCGTTCGGAGACATGATGTAGATAGTT
ATATTTTCCTCCATTTATCTATGGTTCTATAATCAAGTTATCATATATCCTTGTTATCGT
TCTCTAACCATGCCTCCGATCAAGTGGTATGTAAGTCATGATCTATGGATATTCACTCTT
AAGTAAAGATAAATATGAATAATGTCCATGTTTGTAACCACCAATCGCAAGACTGGTCGG
AGGAAGTTTCCTTCGATTAAGGATCGCCCGGGGGGGGGACCATAATGCTATCCTTATTTT
GCTCTTGTTCTTTTTAATTATCTTCTAGTAAGCTCGTAGAAAGGTTGAAGAAAGGTTGAA
ACTAATTATAGAGCGAATTAATACATAGATTTCACTTTCATGCATAATCAGCGACGAAAT

TCCTGCTCAAGATTATCATACTTCCAACAAGAAAAGTTCAGAAATTTCAAATGGTGGGTT
GATGAGATGGATGTAG

>EucaDof21_Eucalyptus_grandis_CDS sequence_forward
ATGTTCGCACCCTCGAGCGACCAAGCGGCGCTGTTCCAGCGGTGGGGGACCTCAAGCCAC
GGCGGCGTCCCTGCCGCGGCGCCAAGCTGCCCCCGCTGCGCCTCCAGCCACACCAAGTTC
TGTTACTACAATAACTACAGCCTCTCTCAGCCCCGATACTTCTGCAAGTCCTGCCGCCGC
GACGAAATTCCTGCTCAAGATTATCATACTTCCAACAAGAAAAGTTCAGAAATTTCAAAT
GGTGGGTTGATGAGATGGATGTAG

>EucaDof21_Eucalyptus_grandis_protein sequence
MFAPSSDQAALFQRWGTSSHGGVPAAAPSCPRCASSHTKFCYYNNYSLSQPRYFCKSCRRR
NSCSRLSYFQQEKFRNFKWWVDEMDV+

>EucaDof22_Eucalyptus_grandis_genomic sequence_reverse
ATGGACACCGCTCAGTGGCCACAGGTAAGATAGATCCATCCATGTCCCTTCTCTCACTCG
ATCTGCATCATCTCCTTCCTCTCTCTCTCTTCTTTTCCTTCTTGATGATTTGGCTGCT
AATAAAGGGTAAAGTGATGCTGTTCTTTGAAGATGAAAGACGATCACTTCTTTTCTTTTC
CTCCTCTCTCACCTCACATAATAGCTGAGTGCAAAGTCTTTACTCTCCTTTGAATTCGGT
CGTGCCTGTTATTGATAGTGTTTTCCACTTTCCCTTCTCCTTTTCAATTACAGTCACCGA
GAAAAAAGAAAAAGGAGAGAGAGAGAGAGAGAGGGAGAGAGAGAGAGACGGTGCATTT
TTATTTTCTGGGCTGATGCCAACTTTGAAAAAGTGAAAAGTGGGCTGTTTTTGCTTTGTT
TGTTGGTCTCGCTTTGTTGGTTTGTGTTCACCTTCTTCCTGTTTGCTTTCTCAGATTTTC
AGGGTTGATCTTCCCTTGTGTTTTGATCTTCATCTGCTCATGCTGCTACGGTTTCTTGAT
GCCTCTTTCATTTCCCTTTTTTCGGCCTTTTTTTCTTTAGGAGATTGTGGTGAAACCAAT
AGAAGATATAGTCACAAGCACATGCACGGCAGCAGCCACCCCAAAGCCATCATCATCTTC
GGTTTCGGAGAGGAAGCCAAGGCCGCAAAAGGAACAAGCCCTCAACTGCCCAAGATGCAA
CTCCACCAACACCAAGTTCTGCTACTACAACAACTACAGCCTCACCCAGCCCAGGTACTT
CTGCAAGACCTGCAGGAGGTACTGGACTGACGGTGGGTCCCTCAGGAACATCCCTGTGGG
CGGTGGCTCCAGGAAGAACAAGAGATCATCGTCCTCAGCATCATCTTCTTCCTCGTCTTT
CAATTCCTCATCCAAGAAGCTGCCTGATCTGATCAGCACACCTGCGTCGAACCCTAACAA
CAAAGTGACACTTCATGAAGGCCAAGATCTGAACCTGGCCTTCCCAAACCCTCACCATCA
TGATTTCAAGTCCATCTCTGAGCTGGTCCAAGTGCCAAGCCTTGAGGCCAGCAAGAACCA
CCATATCTCTGCTAACTCTAGCTCCGCCGGTGCTTCCATGGCACCGCCGCAGCTCTCGGC
TCTGGAGCTGCTCTCGGGGATAACGTCGAGAGGTAGTTTCAGTTCGTTCATGTCGATGCC
CGTCCACGATCCAGGCTCCGTCTACACGCCTGGGCTGTTCGCACTGCCGGACTTCAAGCC
GACCCTCAATTTCTCTCTTGATGGCCTTGGGAGTGGGGGTTATAGGAGCCTCCCTAGTGT
GCAAGAGGGTGGCACTAATGGAGGGAGGCTCTTGTTTCCCTTTGAAGATTTGAAGCCAGT
CTCAAGCACGAGTGATATGGAGCAGAATAGAGGAGATCAAGGAGACTCCAATGGATACTG
GAGTGGGATGTTGGGTGGCGGATCGTGGTA

>EucaDof22_Eucalyptus_grandis_CDS sequence_reverse
ATGGACACCGCTCAGTGGCCACAGGAGATTGTGGTGAAACCAATAGAAGATATAGTCACA
AGCACATGCACGGCAGCAGCCACCCCAAAGCCATCATCATCTTCGGTTTCGGAGAGGAAG
CCAAGGCCGCAAAAGGAACAAGCCCTCAACTGCCCAAGATGCAACTCCACCAACACCAAG
TTCTGCTACTACAACAACTACAGCCTCACCCAGCCCAGGTACTTCTGCAAGACCTGCAGG
AGGTACTGGACTGACGGTGGGTCCCTCAGGAACATCCCTGTGGGCGGTGGCTCCAGGAAG
AACAAGAGATCATCGTCCTCAGCATCATCTTCTTCCTCGTCTTTCAATTCCTCATCCAAG
AAGCTGCCTGATCTGATCAGCACACCTGCGTCGAACCCTAACAACAAAGTGACACTTCAT
GAAGGCCAAGATCTGAACCTGGCCTTCCCAAACCCTCACCATCATGATTTCAAGTCCATC
TCTGAGCTGGTCCAAGTGCCAAGCCTTGAGGCCAGCAAGAACCACCATATCTCTGCTAAC
TCTAGCTCCGCCGGTGCTTCCATGGCACCGCCGCAGCTCTCGGCTCTGGAGCTGCTCTCG
GGGATAACGTCGAGAGGTAGTTTCAGTTCGTTCATGTCGATGCCCGTCCACGATCCAGGC
TCCGTCTACACGCCTGGGCTGTTCGCACTGCCGGACTTCAAGCCGACCCTCAATTTCTCT
CTTGATGGCCTTGGGAGTGGGGGTTATAGGAGCCTCCCTAGTGTGCAAGAGGGTGGCACT
AATGGAGGGAGGCTCTTGTTTCCCTTTGAAGATTTGAAGCCAGTCTCAAGCACGAGTGAT
ATGGAGCAGAATAGAGGAGATCAAGGAGACTCCAATGGATACTGGAGTGGGATGTTGGGT
GGCGGATCGTGGTA

>EucaDof22_Eucalyptus_grandis_protein sequence
MDTAQWPQEIVVKPIEDIVTSTCTAAATPKPSSSSVSERKPRPQKEQALNCPRCNSTNTK
FCYYNNYSLTQPRYFCKTCRRYWTDGGSLRNIPVGGGSRKNKRSSSSASSSSSSFNSSSK
KLPDLISTPASNPNNKVTLHEGQDLNLAFPNPHHHDFKSISELVQVPSLEASKNHHISAN
SSSAGASMAPPQLSALELLSGITSRGSFSSFMSMPVHDPGSVYTPGLFALPDFKPTLNFS
LDGLGSGGYRSLPSVQEGGTNGGRLLFPFEDLKPVSSTSDMEQNRGDQGDSNGYWSGMLG
GGSW#

Additional file 2

Gene	Primer Forward				Primer Reverse				Amplicon Size
	Sequence	Lenght	Tm	%GC	Sequence	Lenght	Tm	%GC	
RibL23A	*AAGGACCCTGAAGAAGGACA*	*20*	*55.5°C*	*50.0*	*CCTCAATCTTCTTCATCGCA*	*20*	*52.3°C*	*45.0*	*128*
Eucons08	*TCCAATCCGAGTCGCTGTCATTGT*	*24*	*60.4°C*	*50.0*	*TGATGAGCCTCTCTGGTTTGACCT*	*24*	*59.8°C*	*50.0*	*152*
Eucons21	*AGAGGTGAAATTCCAGAAGCCCGT*	*24*	*60.4°C*	*50.0*	CTTCCCTTTGGCTTCCGCCAATTA	*24*	*59.8°C*	*50.0*	155
EucaDof3	*AGAAAGTGCCAACGGTACTGGACT*	*24*	*60.0°C*	*50.0*	*ACCAGCACAATCAGCTCGAAGAGA*	*24*	*60.0°C*	*50.0*	153
EucaDof4	*GCATGCAATAATAGGGCCATGTTGG*	*25*	*58.6°C*	*48.0*	*GTTGTTCGTAATTACACCTTGGGC*	*23*	*57.7°C*	*45.8*	*81*
EucaDof6	*TCGATCAAGAACAGGGCGATTCGT*	*24*	*60.1°C*	*50.0*	*GCAGAACTTGGTGTCCATGCTGTT*	*24*	*59.8°C*	*50.0*	*110*
EucaDof10	*TAGCCAACATACCCATGCCAGA*	*22*	*58.1°C*	*50.0*	*TGCAGAAGTGCCTTGGTTGAGAGA*	*24*	*60.2°C*	*50.0*	*114*
EucaDof12	*TCAGGAAATGTCCGGTTCGCTTGA*	*24*	*60.4°C*	*50.0*	*AGCAGAACTTGGTGTTGGTGGAGT*	*24*	*60.5°C*	*50.0*	*134*
EucaDof13	*ATGCAAGATATCCACTCGATCCCAGG*	*26*	*59.8°C*	*50.0*	*AGCAGAACTTGGTGTTGAGCGAGT*	*24*	*60.5°C*	*50.0*	*186*
EucaDof17	*TTAGCCACCACCTTCCAAACCCTA*	*24*	*59.8°C*	*50.0*	*AGGAAACCCAGCTGACGCTGAATA*	*24*	*60.1°C*	*50.0*	*158*
EucaDof21	*CACCAGCAGTTTGCAAGAAGGGTT*	*25*	*60.0°C*	*50.0*	*AGCTAGATGCAAGCAGAGAAGCGA*	*24*	*60.0°C*	*50.0*	*187*
EucaDof22	*ACCTGCGTCGAACCCTAACAACAA*	*24*	*60.5°C*	*50.0*	*AGATATGGTGGTTCTTGCTGGCCT*	*24*	*60.1°C*	*50.0*	*149*

CAPÍTULO III

Evaluation of Putative Reference Genes For Gene Exepression Normalization in *Eucalyptus* species by Microarrays and RT-qPCR

(manuscrito a ser submetido à *BMC Plant Biology*)

Evaluation of putative reference genes for gene expression normalization in *Eucalyptus* species by microarrays and RT-qPCR

Luisa A de Oliveira[1]*, Michèle C Breton[1]*, Fernanda M Bastolla[1]*, Sandro S Camargo[2], Rogério Margis[1], Jeverson Frazzon[1,3], Giancarlo Pasquali[1§]

[1]Graduate Program in Cell and Molecular Biology, Centro de Biotecnologia, Universidade Federal do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, Brazil

[2]Universidade Federal do Pampa, Campus Bagé, Travessa 45, 1650, sala 2107, Bairro Malafaia, CEP 96413-170, Bagé, RS, Brazil

[4]Instituto de Ciência e Tecnologia de Alimentos, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil

[§]Corresponding author: Centro de Biotecnologia, Universidade Federal do Rio Grande do Sul, Av. Bento Gonçalves, 9500, Bairro Agronomia, Caixa Postal: 15053, CEP: 91501-970, Porto Alegre, Rio Grande do Sul, Brazil

*These authors contributed equally to this work

Email addresses:

LAO: lu_abruzzi@yahoo.com.br

MCB: mbreton@uol.com.br

FMB: nandabastolla@yahoo.com.br

SSC: sandro.camargo@unipampa.edu.br

RM: rogerio.margis@ufrgs.br

JF: jeverson.frazzon@ufrgs.br

GP: pasquali@cbiot.ufrgs.br

Abstract

Background

The genus *Eucalyptus* is one of the main sources of wood worldwide and is the most widely used tree species in industrial plantations. Gene expression analysis is increasingly important in biological research, with reverse transcription-quantitative PCR (RT-qPCR) becoming the method of choice for high-throughput and accurate expression profiling of selected genes. Given the increased sensitivity, reproducibility and large dynamic range of this methodology, the requirements for a proper internal reference gene for relative expression normalization have become increasingly stringent. Given the increasing interest in the functional genomics of *Eucalyptus* species, we sought to identify and experimentally verify a set of 15 genes that can be used as a constitutive reference gene for this plant. The present study aimed at identifying suitable reference genes for normalization of gene expression associated with flower, leaf and xylem of six species of *Eucalyptus*.

Results

To provide reliable reference genes in *Eucalyptus sp.*, we surveyed a set of 50 genes selected from microarray of *Eucalyptus grandis* leaves and xylem, and *Eucalyptus globulus* xylem that showed the smallest expression variation when analyzed by two methods, Standard Deviation Microarray Analysis and the Significance Analysis of Microarray. We focused on the most constitutive genes determined by *in silico* analysis of microarray data and further performed the experimental analysis using RT-qPCR for six *Eucalyptus* species and three different organs/tissues. By the use of two distinct algorithms, *geNorm* and *NormFinder*, we have assessed gene expression stability of eight candidate new reference genes. Classic housekeeping, reference genes like those encoding G3PDH, histone H2B, RibL23a, and tubulin were also included in the analysis. The stability profiles of these genes determined with the two different algorithms were in very good agreement.

Conclusions

The normalization strategy is a prerequisite for accurate RT-PCR expression profiling, which, among other things, opens up the possibility of studying the biological relevance of small expression differences, in different tissues in *Eucalyptus* species. Based on the results of *geNorm* and *NormFinder*, the expression of *Eucons04*, *Eucons08* and *Eucons21* genes were the most stable through all organs and species studied in this work, making these genes suitable references for expression studies. These finding will enable more accurate and reliable normalization of RT-qPCR results for gene expression studies in different organs in *Eucalyptus* species economically important.

Background

The genus *Eucalyptus*, with more than 700 species, is one of the main sources of hardwood worldwide and the most widely employed tree in industrial-oriented plantations. Many *Eucalyptus* species are renowned for their fast growth rate, the straight shape of trunks, valuable wood properties, wide adaptability to soils and climates, resistance to biotic stresses, and ease of management through coppicing, seed or clonal propagation. Especially in Brazil, Chile, South Africa, Portugal and India, *Eucalyptus* timber is widely used for cellulose pulp and paper production. Despite the high wood productivity of *Eucalyptus* plantations, reaching 45-60 $m^3.ha^{-1}.year^{-1}$, the increasing demand for cellulose pulp has resulted in wood shortages in recent years [226-228]. Hence efforts in many fields of research are being made to improve forest productivity including molecular approaches like whole genome sequencing and high-throughput analysis of gene expression. With such objectives in mind, The *Eucalyptus* Genome Network (EUCAGEN) was created (http://www.ieugc.up.ac.za), representing one example of a valuable database platform for genome research in *E. grandis* and other species [229].

With the recent availability of genome and transcriptome data, many efforts are and will be done to assess *Eucalyptus* gene expression with conventional or high-throughput techniques. Independently of the method employed, it is absolutely essential the use of reference genes as internal controls for gene expression measurements. The availability of such validated reference genes for *Eucalyptus* is still scarce.

DNA macro and microarray hybridizations, partial or whole transcriptome sequencing linked to digital transcript counting (RNA-Seq), among other techniques allow the expression analysis of thousands of genes simultaneously, employing differentially labeled RNA or cDNA populations. These techniques have the advantage of speed, high-throughput and a high degree of potential automation compared to conventional quantification methods such as northern blot analysis, ribonuclease protection assays, or competitive RT-PCR [207, 230, 231]. Reverse-transcription followed by real-time, quantitative polymerase

chain reaction (RT-qPCR) is the most sensitive and specific technique commonly used to assess gene expression levels [232]. It allows performing more in-depth studies of smaller sets of genes across many individuals, treatments or cell/tissue types. RT-qPCR is the technique of choice to validate gene expression results derived from the mentioned high-throughput methods [207, 230, 231].

As referred previously, only good internal reference genes will allow confident comparison of gene expression results. Internal control genes are used to normalize mRNA fractions and are often referred to as housekeeping genes which should not vary their expressions during development, among tissues or cells under investigation, or in response to experimental treatments. Most common housekeeping genes employed in plant gene expression studies are those encoding actin [233, 234], tubulin [235, 236], glyceraldehyde-3-phosphate dehydrogenase [237, 238], ribosomal RNA [235, 236, 239], polyubiquitin [233, 240], and elongation factor 1-α (EF1-α) [237, 241]. Many studies make use of these housekeeping genes without proper validation of their presumed stability, based on the assumption that they would be constitutively expressed due to their role in basic cellular processes. Considerable amount of data show that most studied housekeeping genes have expressions that can vary considerably depending on the cell type or experimental condition [242, 243]. With the increased sensitivity, reproducibility and large dynamic range of the RT-qPCR methods, the requirements for proper internal control genes have become increasingly stringent.

In recent years, a large number of reference gene validation attempts have been reported for plants, most of them covering model and crop species like rice [244], *Arabidopsis thaliana* [245], tobacco [235], sugarcane [246], potato [247], *Brachypodium* sp. [248], soybean [249, 250], tomato [251, 252], *Brachiaria* sp. [241], coffee [233], peach [237], wheat [253], chicory [238], cotton [254], and cucumber [255]. Few studies have focused on forest genera such as poplar [256, 257], grape [258], and longan tree [259]. Reference genes for gene expression studies in *Eucalyptus* have been recently presented. Almeida *et al*. [260], working with *E. globulus* microccuttings rooted *in vitro*, have indicated histone H2B and α-tubulin the most suited reference genes during *in vitro* adventitious rooting, in the presence or absence of auxin. Boava *et al*. [261], working with clonal seedlings of

the hybrid plant (*E. grandis* x *E. urophylla*) exposed to biotic (*Puccinia psidii*) or abiotic (acibenzolar-S-methyl) stresses, concluded that genes encoding the eukaryotic elongation factor 2 (eEF2) and ubiquitin were the most stable, and ideal for internal controls. Both works tested a small number of genes (11 and 13, respectively) selected according to literature data concerning other plant systems and experimental conditions.

Given the increasing interest in the functional genomics of *Eucalyptus* and the need of validated reference genes for a broader set of species and experimental conditions, we sought to identify the most stably expressed genes in a set of 21,442 genes assayed by microarray developed to compare stem vascular (xylem) and leaf tissues of *E. grandis* and *E. globulus* adult trees. Best candidate genes were than validated by RT-qPCR in assays with RNAs from xylem and leaves of six *Eucalyptus* species and flowers of *E. grandis*. Seven traditional housekeeping genes most employed in expression studies in plants were also included in our analysis. As a result, genes selected as the least variable among all conditions tested were not yet described in literature. These genes may represent an important molecular tool to accurately analyze the expression of *Eucalyptus* genes in different tissues/organs and in different species via array hybridization or RT-qPCR.

Results

Selection of *Eucalyptus* reference genes via microarray analysis

Data resulted from microarray hybridizations conducted within Project "*Genolyptus*: The Brazilian Research Network on the *Eucalyptus* Genome" (http://genoma.embrapa.br/genoma/genolyptus) were analyzed for selecting the most stably expressed *Eucalyptus* genes. The microarray study was conceived with nine 50mer-oligoprobes covering the length of each one of the 21,442 unique sequences derived from the *Genolyptus* EST dataset. Oligoprobes were synthesized “on-chip” in duplicate, randomly distributed in two blocks of ten identical slides. Leaf blades and vascular (xylem) tissue samples were taken from two *E. grandis* clonal

trees, i.e., derived from the same matrix tree and harboring the same genotype. Two additional xylem samples were collected from two other *E. grandis* clonal trees of a different genotype and from two *E. globulus* clonal trees. Therefore, ten Cy3-labeled cDNA samples and ten identical chips were produced at Roche NimbleGen for the microarray assays, with a total number of 385,956 features per slide.

Most stably expressed genes were mined in the microarray data with the employment of two statistical algorithms named Significance Analysis of Microarrays (SAM; [262] and Standard Deviation Microarray Analysis (SDMA; Camargo et al., Ph.D. thesis), that allows the representation of results in three dimensional (3D) graphics.

The input data to SAM were gene expression measurements from the set of microarray experiments, as well as the response variable from each experiment. According to Tusher *et al*. [262], SAM computes a statistic *d*(i) for each gene "i", measuring the strength of the relationship between gene expression and the response variable. It uses repeated permutations of the data to determine if the expression of any gene is significantly related to the response. The cutoff for significance is determined by a tuning parameter *delta*, chosen by the user based on the false positive rate. One can also choose a fold change parameter, to ensure that called genes change at least a pre-specified amount. In the present study, the value of *delta* was set to 0.2 so that we could mine the genes whose expressions exhibited the lowest variation among the three conditions assessed in the microarrays, i.e., *E. grandis* leaves and xylem and *E. globulus* xylem (Figure 1A). A ranking of 50 genes whose fold change values were approximately equal to one were selected as candidates for reference genes, since they presented the lowest variation of expression among leaves and xylems of *E. grandis* and *E. globulus* (Table 1).

SDMA is a novel and simpler algorithm based on the comparison of average gene expressions in relation to the global average of expressed genes in microarrays and the overall standard deviation, allowing the presentation of results in graphical mode (Camargo et al., Ph.D. thesis). SDMA allowed us to generate a 3D graph that evidenced genes expressed in a position equivalent to their overall average expression among the three conditions analyzed in the microarrays (Figure

1B). The average value of gene expression by SDMA should be as similar as possible to the global average of expression, and the overall standard deviation should tend to zero when the scope of the analysis is the selection of genes whose expressions are stable. Using the same criteria applied in SAM, we selected 50 genes whose standard deviations were close to zero, indicating the similarity between the values of mean and mean global gene expression (Table 1). A SDMA 3D graphic is presented in Figure 1C where the mean expressions of the 50 most stable genes selected under the conditions studied are plotted. Note that points representing selected genes tend to form a straight line, indicating that their means of expression when compared with the global average have a standard deviation tending to zero.

We were pleased to note that the employment of either SDMA or SAM allowed us to identify the same group of 50 genes as the most stably expressed, confirming the robustness of the analysis performed by both algorithms. Nevertheless the ranking of the two methods differed, as presented in Table 1. Since none of the sequences selected presented molecular or biochemical identities similar to previously described *Eucalyptus* genes or proteins, we named them *Eucons01* to *Eucons50*, according to the ranking generated by SDMA, as stated in the first column of Table 1.

Selected sequences were annotated using BLASTx [65] against the available non-redundant protein sequences (nr) and their functional categories were determined by the Blast2GO software [64]. Although some sequences exhibited Expected (E) values too high for a confident annotation, approximately half of them (48%) showed similarity to known proteins. The other half of the sequences corresponded to hypothetical proteins (10%) or returned a "no hit" (42%) result (Table 1). The gene ontology analysis of the 50 selected genes by Blast2GO allowed the functional classification of 35 (70%) of the sequences, as represented in Figure 2. Most of the sequences were classified in functions related to cellular (12) or metabolic (12) processes, among six other functional categories. The remaining 15 (30%) sequences were classified as "no hit", and were not represented in Figure 2.

In order to further validate results by RT-qPCR, we selected the five candidate genes with the littlest variation in expression and whose annotation

matched a known plant protein according to SDMA (*Eucons01*, *04*, *06*, *07* and *08*) and SAM (*Eucons15*, *21*, *27*, *32* and *43*). Selected genes are highlighted in gray in Table 1.

Validation of *Eucalyptus* reference genes by RT-qPCR

In order to check their true stability in expression, primers for RT-qPCR validation of the ten *Eucalyptus* sequences elected as reference genes were designed and are presented in Table 2. Besides them, we also designed primers for five genes traditionally employed as references, based on their housekeeping function, including a SAND family protein [245], glyceraldehyde-3-phosphate dehydrogenase (G3PDH) [263], histone H2B, ribosomal protein L23A, and tubulin (TUA2) [207], as presented in Table 3. Reference genes previously recommended for the analysis of gene expression during *E. globulus* rooting *in vitro* and named *Euc10* and *Euc12* were also evaluated [260] and primers employed in RT-qPCR are also presented in Table 3.

Total RNA samples were prepared from six *Eucalyptus* species, distributed as follows: flower, leaf and xylem of *E. grandis*, leaf and xylem of *E. dunnii*, *E. pellita*, *E. saligna* and *E. urophyla*, and xylem of *E. globulus*. RT-qPCR evaluations were conducted with biological duplicates and experimental quadruplicates. Results were analyzed using softwares *geNorm* 3.5 [264] and *NormFinder* [265] in order to generate comparable rankings of genes based on their stability of expression. The Cq data collected for all samples were transformed to relative quantities using the $2^{-\Delta\Delta Ct}$ method developed by Livak and Schmittgen [194]. We did not succeed in obtaining satisfactory dissociation curves after RT-qPCR with primers designed for *Eucons01* and *Eucons15* (data not show) and both candidate genes were discarded from the analysis.

With *geNorm*, the average expression stability (*M* value) of all genes was first calculated. The *M* values are defined as the mean variation of a certain gene related to all of the others. The *geNorm* software recommends an *M* value below the threshold of 1.5 in

order to identify genes with stable expression, although 0.5 has been used as threshold limit by many authors [260, 266-269]. As shown in Figure 3A, all 15 candidate genes examined showed a very high stability of expression with thresholds lower than 0.12, independently of tissues/organs evaluated. According to the *geNorm* analysis, *Eucons04*, *Eucons08* and *Eucons21* are the most stably expressed ones and should be considered as best reference genes for RT-qPCR normalizations.

In order to evaluate the optimal number of reference genes for reliable normalization, *geNorm* allows to calculate the pairwise variation V*n*/V*n*+1 between the sequential ranked normalization factors NF*n* and NF*n*+1 to determine the effect of adding the next reference gene in normalization. The normalization factor is calculated based on the geometric average among the two most stable gene relative quantities and the stepwise inclusion of the other genes in the order of their expression stability. A large pairwise variation implies that the added reference gene has a significant effect on normalization and should be included for calculation of a reliable normalization factor. Considering the *cut-off* value of 0.15, below which the inclusion of an additional reference gene is not necessary [264], the use of the two most stably expressed genes *Eucons08* and *Eucons21* was sufficient for accurate normalization ($V_{2/3}$ = 0.006) in all organs studied (flower, leaves and xylem) from the six *Eucalyptus* species (Figure 3B). The same applies when analyzing xylem and leaves separately, with a $V_{2/3}$ = 0.006 for *Eucons04* and RibL23a genes (leaves) and $V_{2/3}$ = 0.004 for *Eucons06* and *Eucons08* genes (xylem; data not show).

In addition to *geNorm*, the expression stability of candidate reference genes assayed by RT-qPCR was also analyzed with the *NormFinder* software. This program takes into account the intra- and inter-group variations for normalization factor calculation and the results are not affected by occasionally co-regulated genes. The best candidate will be the one with the inter-group variation as close to zero as possible and, at the same time, having the smallest error bar possible. Hence values are inversely correlated to gene expression stability, which avoids artificial selection of coregulated genes [259].

According to the *NormFinder* analysis of gene expressions in leaves, xylem tissues and among species, the stability values of the 15 genes studied were lower than 0.138, with error bars no greater than 0.044 (Figure 3C; Table 4). When

we analyzed the gene expressions in all tissues/organs and species, the stability value was in the range between 0.017 and 0.106, proving again that all genes elected are good references for RT-qPCRs studies in *Eucalyptus*. The ranking of the genes and their respective stability values are shown in Table 4. According to the *NormFinder* analysis and in agreement with the results of *geNorm*, the three most stable genes were *Eucons04*, *Eucons08* and *Eucons21* when considering all tissues/organs and species. When the expressions in leaves are separately considered, the stability values were in the range between 0.008 and 0.086, and the three most stable genes in these organs were *Eucons04*, *Eucons08* and *Eucons32*. In xylem vascular tissues, the stability values were in the range of 0.01 to 0.138, and genes *Eucons27*, *Eucons07* and *Eucons06* were the most stable (Table 4). The algorithm ranked *Eucons04* as the most stably expressed gene in all samples regardless of whether the samples were collected into one main group or divided in two groups. Nonetheless, just one housekeeping gene is determined to all samples using *NormFinder* when no groups are defined. So, a different group was created to analyze the most stably couple (Table 4). When leaves and xylem were tested as different groups the stability values were in the range between 0.011 and 0.094. *Eucons04* exhibited the lower stability value. *NormFinder* identified *Eucons04* and *Eucons08* as the most appropriate combination of genes, showing a stability value of 0.009.

Confirming the stability of *Eucalyptus* reference genes via *dxr* differential gene expression

Terpenoids are all derived from two common precursors, isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP). In higher plants, IPP and DMAPP are synthesized through two distinct pathways in separate cellular compartments, the cytosolic mevalonate (MVA) pathway and the 2-C-methyl-D-erythritol 4-phosphate (MEP) pathway that takes place in plastids. The MEP pathway, through which diterpenes are synthesized, has two important initial steps: (i) the formation of 1-deoxy-D-xylulose 5-phosphate (DXP) from piruvate and

glyceraldehyde 3-phosphate through the action of the DXP synthase (DXS), followed by the conversion of DXP into MEP by the action of the DXP reductoisomerase (DXR). As DXS and DXR are key enzymes catalyzing the two committed steps for isoprenoid biosynthesis, genes coding for DXS and DXR may play important roles in controlling the plastidic synthesis of isoprenoids and the downstream diterpene products [270].

It is known that the expression patterns of the DXR enzyme and its encoding gene vary quite consistently according to the plant organ being assessed. This enzyme shows increased expression in leaves of *A. thaliana* and a number of other plants, but decreased levels were reported in stems and flowers [271-273]. In order to confirm the constitutive expression of the three best *Eucalyptus* genes selected as references (*Eucons04*, *Eucons08* and *Eucons21*), we tested them by normalizing the patterns of *dxr* gene expression in *Eucalyptus* and compared the results with those normalized by the traditional reference genes RibL23a and G3PDH. Therefore the *dxr* and reference gene expressions were measured by RT-qPCR in the same set of tissues/organs and *Eucalyptus* species previously tested. *Dxr* expression values were than normalized against the expression values of two reference genes, as shown in Figure 4.

In order to allow comparisons among reference genes, the average value of the pairwise reference gene relative expressions in the different organs/tissues tested was set to one (1) and taken to normalize the *dxr* relative expression. As expected, steady-state mRNA levels for the *dxr* gene were much higher in leaves followed by flowers, with lower values observed in xylem tissues of *E. grandis* (Figure 4B). As shown in Figure 4, the pairwise combination of *Eucons04*, *Eucons08* or *Eucons21* allowed more confident results than the RibL23/G3PDH pair. The relative expression of the *dxr* gene was much less variable when normalized with *Eucons* genes and, most importantly, much more concordant if compared to results normalized by the RibL23a/G3PDH pair. This is more evident in Figure 4B where no statistical difference was observed in *dxr* relative expression values among xylems from *E. grandis*, *E. globulus* and *E. pellita*, but was quite different when normalized with any two of the *Eucons* genes. Essentially the same conclusions were assumed by the analysis of *dxr* relative expression obtained with leaves and xylem tissues from *E. dunnii*, *E. pellita*, *E. saligna* and *E. urophylla*, not shown in Figure 4.

Discussion

Real-time PCR and cDNA microarray measurements are highly reproducible tecniques [274-276]. However, in comparison to classical reverse transcription-polymerase chain reaction (RT-PCR), the main advantages of real-time PCR are its higher sensitivity, specificity, and broad quantification range of up to seven orders of magnitude [244, 277, 278], besides being a great aid to study expression in genes whose transcript levels are known to be very low [254, 279]. Real-time PCR analysis has become the most common method for validating the whole-genome microarray data or a smaller set of genes, and molecular diagnostics [244, 280-282].

Accurate normalization is an absolute prerequisite for correct measurement of gene expression, and most commonly used normalization strategy involves standardization to a single constitutively expressed control gene [265]. Therefore, the reference gene ideal should exhibit invariable expression levels among all different cell types, organs, developmental stages or treatments that are submitted [254, 265].

However, it has become clear that no single gene is constitutively expressed in all cell types and under all experimental conditions [252, 265, 283]. As it has been shown that the expression of such housekeeping genes, although constant in some experimental conditions, can vary considerably in other cases [207, 247, 284] implying the expression stability of the intended control gene has to be verified before each experiment. Normalization with multiple reference genes is becoming the golden standard for the technique, but reports that identify such genes in plant research are limited [241, 244, 248, 258, 285], especially for woody species.

In the present work, we evaluated by RT-qPCR the expression stability of 15 candidate genes in different organs (leaves and flowers) and vascular tissues (xylem) derived from six species of *Eucalyptus*. Analyzed genes were selected based on microarray data of 21,442 *Eucalyptus* genes and previous studies of RT-qPCR normalization in plants such as *A. thaliana* and *Populus* sp. [235-238, 284, 286, 287]. Among the candidates tested, classic housekeeping genes derived from the works

were included like those encoding tubulin, histone H2B, ribonucleoprotein L23A and G3PDH.

The stability of the tubulin gene has been often used to normalize RT-qPCR expression data [256, 288]. Analysis of the *Eucalyptus* microarray and RT-qPCR data revealed that, indeed, it has a quite stable expression. Nevertheless, this gene is far from being the best reference for *Eucalyptus* among those tested (Figure 3 and Table 4). Tubulin has been shown to be a suitable normalization gene during development in orobranche [288], but it was apparently unstable during seedling development in *A. thaliana* [284, 289, 290], differenttissue, biotic and abiotic stress in potato [247] and cucumber [255]. Similar results were obtained by Artico *et al* [254] and Silveira *et al*. [241] respectively with *G. hirsutum* and *B. brizantha*.

Genes encoding G3PDH, histone H2B and ribonucleoprotein L23A are the most employed and tested housekeeping genes in plants [207, 231, 233, 243, 246, 260, 284]. According to our RT-qPCR results and data analysis by geNorm and NormFinder, all these three housekeeping genes showed quite consistent stability in expression in *Eucalyptus*, especially RibL23A (Figure 3 and Table 4). Nevertheless, the employment of the pair G3PDH/RibL23A to normalize the expression of the isoprenoid biosynthetic gene *dxr* (Figure 4) proved that, at least together, these genes are not suited as references for *Eucalyptus* gene expression. According Kim *et al.* [291], the relative expression of G3PDH based on *O. sativa* Dasan cultivar varied up to two-fold. The analysis results of RT-qPCR of *Brachypodium distachyon* showed that the G3PDH gene was stably expressed under various abiotic stress conditions and not discernibly affected by diverse growth hormones, although it exhibited less stability in some plant tissues [248]. In tomato, G3PDH showed the top-ranked by geNorm, with phosphoglycerate kinase gene (PGK) under light stress. However, were poorly ranked in the cold stress and growth conditions of nitrogen starvation experiments [292]. Similar results were obtained with peach, where G3PDH was not among the best reference genes between experimental groups. Reasons for those discrepancies may be that GAPDH not only acts as a component of the glycolytic pathway but also takes part in other processes as well [237]. Therefore, the expression profile of G3PDH might fluctuate according to the corresponding experimental conditions.

Ribosomal protein genes are often viewed as a homogeous collection of housekeeping genes with little attention aid to the diversity of this large group. This gene family shown to have extraribosomal functions [293, 294] they have also been routinely identified individually or as part of small groups in screens for genes specifically induced or repressed in particular tissues during different stages of development as tuber development [295] and root development [296]; or during stresses, as genotoxic stress [297], cold stress [298, 299], heat stress [300], and depletion of endoplasmic reticulum calcium stores [301]. Volkov *et al.* [284] evaluated tissue-specific changes in the mRNA levels of RibL23A gene measured in different organs of *A. thaliana*. Compared to leaves, the level of RibL23A mRNA was increased in flowers and reduced in stems and siliques. These observations are in accordance with the idea that ribosomal protein genes in plants are transcriptionally up-regulated in actively growing tissue and down-regulated in metabolically inactive tissues [302, 303].

The sequences called *Euc10* and *EuC12* showed similarity with the protein sequence corresponding to a hypothetical protein [304] and a putative RNA binding protein [260, 304] from *A. thaliana*, respectively. These genes showed expression levels were more stable in Bastolla [304] evaluation for *E. grandis* and *E. globulus* leaf and xylem tissue. In the Almeida *et al.* [260] analysis, *Euc12* was identified as the third most stable reference gene during *in vitro* adventitious rooting in *E. globulus*. Although these genes show stable expression levels for these tissues and specifics conditions, its expression profile was very variable for all *Eucalyptus* species evaluated in this study.

Eucalyptus microarray data of 21,442 genes allowed us to identify the genes whose expressions were the most invariable among xylems of *E. grandis* and *E. globulus* and leaves of *E. grandis* (Table 1). RT-qPCR analysis of ten selected genes among the most stably expressed proved that these genes were indeed very reliable references for the normalization of gene expression in different *Eucalyptus* organs and tissues, especially those named *Eucons04*, *Eucons08* and *Eucons21*. Analysis of the function of the putative encoded proteins revealed that they may also belong to the so called housekeeping class of genes.

The *Eucons04* gene encodes a protein highly similar to putative cyclin-dependent protein kinases (CDK) such as the *R. communis* CDK8 (E-value of 2e-49; Table 1) and CDKs from *Arabidopsis* [305]. These type of proteins are able to phosphorylate protein-target amino acids in different metabolic pathways and, most notably, in cell cycle control [306-308].

Eucons08 is similar (E-value of 1e-25; Table 1) to *R. communis* and *A. thaliana* genes possibly encoding the transcription elongation factor s-II (TFIIS). The product of this gene facilitates RNA polymerase II read-through of various blocks to transcript elongation by strongly enhancing the intrinsic RNA nuclease activity of RNA polymerase II [309].

Eucons21 encodes a protein with significant sequence similarity to a putative *R. communis* aspartyl-tRNA synthetase (E-value of 2e-42; Table 1). Aminoacyl-tRNA synthetases catalyze the addition of amino acids to their cognate tRNAs. In the case of aspartyl-tRNA synthetase, the aminoacid bound to tRNAs is aspartate. In plants, all aminoacyl-tRNA synthetases are nuclear-encoded and are post-translationally targeted to the compartments where protein synthesis takes place, i.e. cytoplasm, mitochondria or plastids [310].

According to the analysis of the RT-qPCR data performed with software *NormFinder*, *Eucons04* and *Eucons08* are the best reference genes pairwisely when assessing test gene expression in leaves alone, or in leaves along with xylem tissues. If xylem tissues are solely analyzed, *Eucons07* and *Eucons27* would be best references (Table 4). *Eucons07* encodes a similar protein to a ABC transporter family protein ATNAP10 from *A. lyrata* (E-value of 1e-28; Table 1) and Eucons27 enodes a similar putative peroxisome biogenesis factor from *R. communis* (E-value of 3e-63; Table 1).

In a *Cyclamen persicum* study, the average expression stability value (M) generated by geNorm displayed *ABC transporter ATPase* one of most stable transcript abundance, was indicated as a normalizing gene for studies in the experimental system in cell cultures in this specie [311]. Analyzing the peroxisomal genes expression profiles in *A. thaliana*, Kamada *et al.* [312] evaluated genes identified as peroxisomal biogenesis showing that these factors are expressed in all

plant organs, which suggests that they play a role in metabolic pathways of unidentified plant peroxisomes and may have a constitutive expessionin plant.

Conclusions

Based on microarray expression analysis of more than 20,000 *Eucalyptus* genes, we identified 50 new constitutively expressed genes suited as references for the normalization of *Eucalyptus* gene expression data. We proved by RT-qPCR that representatives of these reference genes are indeed vary stable in different organs/tissues and species of *Eucalyptus*. We therefore proved our assumption about the level of expression by experimental analysis and not simply by annotation based on a housekeeping function. Considering that two statistical programs allowed us to reach similar interpretations of the results, and that potential discrepancies should be expected, the good agreement of our results with these independent approaches strongly suggests that *Eucons04*, *Eucons08* and *Eucons21* can be regarded as suitable reference genes for normalization in *Eucalyptus* species. In summary, these findings provide useful tools for normalization of RT-qPCR experiments and will enable more accurate and reliable gene expression studies related to functional genomics in *Eucalyptus*.

Methods

Plant Material

Field grown *Eucalyptus grandis* and *Eucalyptus globulus* trees planted in 2002 at Hortoflorestal Barba Negra (Aracruz Celulose S.A.) in Barra do Ribeiro, RS, Brazil were used in this study. Samples were collected from clonal trees. Xylem and leaves of the five *Eucalyptus* species were collected in 2008: *E. dunnii*, *E. grandis*, *E.*

pellita, *E. saligna* and *E. urophylla*. *E. grandis* flowers and *E. globulus* xylem were also collected under the same conditions. Harvested organs/tissues were immediately frozen in liquid nitrogen and stored at -80 °C until further analysis.

RNA extraction

Total RNA was extract using the PureLink Plant RNA Purification (Invitrogen) according to the manufacturer's instructions for small scale RNA isolation. About 40 μg of total RNA was sent to NimbleGen Systems Inc. (Reykjavik, Island) for cDNA synthesis and microarray hybridizations.

Oligonucleotide microarray hybridization

In order to have an evaluation of gene expression patterns in leaf and vascular tissues of two different *Eucalyptus* species, *E. grandis* and *E. globulus*, microarray analysis including all unique EST sequences was carried out. Twenty microgram of each RNA sample was dried out and sent to NimbleGen Systems Inc. Nine 50-mer probes were designed and synthesized in duplicates from the 21,442 ESTs of the Genolyptus Project at NimbleGen. In total, 8 samples, being 2 replicates for 2 replicas of 2 different genotypes. Synthesis of Cy3-labeled cDNAs, hybridizations, washings, scanning and preliminary analyses were carried out by NimbleGen, following a standard expression design. Normalized gene expression values for each feature were obtained taking into account the hybridization values for all 9 probes per gene in duplicates.

Data processing and statistical analysis

Microarray expression data were normalized into log2 intensity values. Afterwards, we have done three distinct analyses. In the first one, we have compared hybridizations from *E. grandis* leaf and xylem. In the second one, we have compared *E. grandis* xylem and *E. globulus* xylem. In both previous analyses, the aim was to find the most similarly and the most differentially expressed genes. In the

third analysis, we have looked for the most similarly expressed genes in hybridizations from three tissues.

In each analysis, data was mean-centered as follows. A reference set is generated by averaging the expression of each gene over all hybridizations. Each hybridization data was subtracted from the reference data set, generating new mean-centered data. In the next step, the "relative difference" in gene expression was computed. The relative difference score was used to identify the most similarly and the most differentially expressed genes. We have performed the Two Class Unpaired Significance Analysis of Microarrays (SAM; [262]) when comparing two tissues, and a Multiclass SAM when comparing three organs. In order to perform the experiments we have used SAM Version 3.09 and R 2.9.2 tools and SDMA V1.0 tool.

RT-qPCR

To assess the reproducibility of the method used in this study, biological and technical replicates from xylem and leaf RNA samples were submitted to RT-qPCR. A subset of twenty genes was used to verify the microarray results. Primers pairs were designed using PrimerQuest (http://www.idtdna.com/Scitools/Applications/Primerquest) and are listed in Table 2. The relative transcript abundance was detected by SYBR Green and PCRs were carried out in a total volume of 20 µL using the StepOne™ *Real-Time* PCR System (Applied Biosystems). Reaction conditions were: one cycle of denaturation at 95 °C for 10min followed by 40 cycles of 95 °C for 15 sec (denaturation) and 60 °C for 15 sec (annealing and elongation). PCRs were followed by a melting curve program (60 to 95 °C with a heating rate of 0.1 °C per second and a continuous fluorescence measurement). A negative control was run without cDNA template in all assays to assess the overall amplification specificity.

Annotation

The Gene Ontology functional annotation tool Blast2GO [64] was used to assign GO identities and enzyme commission numbers. This tool also enabled

statistical analysis related to over representation of functional categories based on a Fisher Exact statistic methodology.

Authors' contributions

LAO, MCB and FMB participated in the preparation of biological samples, RT-qPCR analysis, statistical analysis, annotation data and helped to draft the manuscript. SSC performed the microarray analysis, statistical analysis, construction of graphs and tables. JF conceived and participated in the coordination of the study. GP conceived the study, participated in its design and coordination helped performing the microarray study. All authors read and approved the final paper.

Acknowledgements

We acknowledge M.Sc. Marta Dalpian Heis for her guidance and review of the statistical tests. We are grateful to the staff of Aracruz Celulose S.A. for providing the biological material and Genolyptus Project members that selected the *E. grandis* and *E. globulus* material. This work was financially supported by *Conselho Nacional de Desenvolvimento Científico e Tecnológico* (CNPq, Brazilian Ministry of Science and Technology) and *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior* (CAPES, Brazilian Ministry of Education).

REFERENCES

226. Foucart C, Paux E, Ladouce N, et al. **Transcript profiling of a xylem vs phloem cDNA subtractive library identifies new genes expressed during xylogenesis in *Eucalyptus*.** *The New phytologist* 2006, **170**:739-52.

227. Grattapaglia D, Kirst M: ***Eucalyptus* applied genomics: from gene sequences to breeding tools.** *The New phytologist* 2008, **179**:911-29.

228. Steane DA, Nicolle D, McKinnon GE, Vaillancourt RE, Potts BM: **Higher-level relationships among the eucalypts are resolved by ITS-sequence data**. *Australian Systematic Botany* 2002, **15**:49-62.

229. Rengel D, San Clemente H, Servant F, et al. **A new genomic resource dedicated to wood formation in *Eucalyptus***. *BMC Plant Biology* 2009, **9**:36.

230. Kim S, Kim T: **Selection of optimal internal controls for gene expression profiling of liver disease.** *BioTechniques* 2003, **35**:456-8, 460.

231. Rajeevan MS, Vernon SD, Taysavang N, Unger ER: **Validation of array-based gene expression profiles by real-time (kinetic) RT-PCR.** *The Journal of molecular diagnostics : JMD* 2001, **3**:26-31.

232. Aerts JL, Gonzales MI, Topalian SL: **Selection of appropriate control genes to assess expression of tumor antigens using real-time RT-PCR.** *BioTechniques* 2004, **36**:84-6, 88, 90-1.

233. Barsalobres-Cavallari CF, Severino FE, Maluf MP, Maia IG: **Identification of suitable internal control genes for expression studies in Coffea arabica under different experimental conditions**. *BMC Molecular Biology* 2009, **10**:1.

234. Bas A, Forsberg G, Hammarström S, Hammarström M-L: **Utility of the housekeeping genes 18S rRNA, beta-actin and glyceraldehyde-3-phosphate-dehydrogenase for normalization in real-time quantitative reverse transcriptase-polymerase chain reaction analysis of gene expression in human T lymphocytes.** *Scandinavian journal of immunology* 2004, **59**:566-73.

235. Schmidt GW, Delaney SK: **Stable internal reference genes for normalization of real-time RT-PCR in tobacco (*Nicotiana tabacum*) during development and abiotic stress**. *Molecular Genetics and Genomics* 2010, **283**:233-41.

236. Yang Y, Hou S, Cui G, et al. **Characterization of reference genes for quantitative real-time PCR analysis in various tissues of *Salvia miltiorrhiza***. *Molecular Biology Reports* 2010, **37**:507-13.

237. Tong Z, Gao Z, Wang F, Zhou J, Zhang Z: **Selection of reliable reference genes for gene expression studies in peach using real-time PCR**. *BMC Molecular Biology* 2009, **10**:71.

238. Maroufi A, Van Bockstaele E, De Loose M: **Validation of reference genes for gene expression analysis in chicory (*Cichorium intybus*) using quantitative real-time PCR**. *BMC Molecular Biology* 2010, **11**:15.

239. Guénin S, Mauriat M, Pelloux J, et al. **Normalization of qRT-PCR data: the necessity of adopting a systematic, experimental conditions-specific, validation of references.** *Journal of experimental botany* 2009, **60**:487-93.

240. Libault M, Thibivilliers S, Bilgin DD, et al. **Identification of four soybean reference genes for gene expression normalization**. *The Plant Genome* 2008, **1**:44-54.

241. Silveira ED, Alves-Ferreira M, Guimarães LA, Silva FR da, Carneiro VTDC: **Selection of reference genes for quantitative real-time PCR expression studies in the apomictic and sexual grass *Brachiaria brizantha***. *BMC Plant Biology* 2009, **9**:84.

242. Hruz T, Wyss M, Docquier M, et al. **RefGenes: identification of reliable and condition specific reference genes for RT-qPCR data normalization**. *BMC Genomics* 2011, **12**:156.

243. Thellin O, Zorzi W, Lakaye B, et al. **Housekeeping genes as internal standards: use and limits**. *Journal of Biotechnology* 1999, **75**:291-295.

244. Jain M, Nijhawan A, Tyagi AK, Khurana JP: **Validation of housekeeping genes as internal control for studying gene expression in rice by quantitative real-time PCR**. *Biochemical and biophysical research communications* 2006, **345**:646-51.

245. Remans T, Smeets K, Opdenakker K, et al. **Normalisation of real-time RT-PCR gene expression measurements in *Arabidopsis thaliana* exposed to increased metal concentrations**. *Planta* 2008, **227**:1343-9.

246. Iskandar HM, Simpson RS, Casu RE, et al. **Comparison of reference genes for quantitative real-time polymerase chain reaction analysis of gene expression in sugarcane**. *Plant Molecular Biology Reporter* 2004, **22**:325-337.

247. Nicot N, Hausman JF, Hoffmann L, Evers D: **Housekeeping gene selection for real-time RT-PCR normalization in potato during biotic and abiotic stress**. *Journal of Experimental Botany* 2005, **56**:2907-2914.

248. Hong S-Y, Seo PJ, Yang M-S, Xiang F, Park C-M: **Exploring valid reference genes for gene expression studies in *Brachypodium distachyon* by real-time PCR.** *BMC plant biology* 2008, **8**:112.

249. Hu R, Fan C, Li H, Zhang Q, Fu Y-F: **Evaluation of putative reference genes for gene expression normalization in soybean by quantitative real-time RT-PCR**. *BMC Molecular Biology* 2009, **10**:93.

250. Jian B, Liu B, Bi Y, et al. **Validation of internal control for gene expression study in soybean by quantitative real-time PCR**. *BMC Molecular Biology* 2008, **9**:59.

251. Coker JS, Davies E: **Selection of candidate housekeeping controls in tomato plants using EST data.** *BioTechniques* 2003, **35**:740-2, 744, 746 passim.

252. Expósito-Rodríguez M, Borges AA, Borges-Pérez A, Pérez JA: **Selection of internal control genes for quantitative real-time RT-PCR studies during tomato development process**. *BMC Plant Biology* 2008, **8**:131.

253. Paolacci AR, Tanzarella OA, Porceddu E, Ciaffi M: **Identification and validation of reference genes for quantitative RT-PCR normalization in wheat.** *BMC molecular biology* 2009, **10**:11.

254. Artico S, Nardeli SM, Brilhante O, Grossi-de-Sa MF, Alves-Ferreira M: **Identification and evaluation of new reference genes in *Gossypium hirsutum* for accurate normalization of real-time quantitative RT-PCR data.** *BMC plant biology* 2010, **10**:49.

255. Wan H, Zhao Z, Qian C, et al. **Selection of appropriate reference genes for gene expression studies by quantitative real-time polymerase chain reaction in cucumber**. *Analytical Biochemistry* 2010, **399**:257-61.

256. Brunner AM, Yakovlev IA, Strauss SH: **Validating internal controls for quantitative plant gene expression studies**. *BMC Plant Biology* 2004, **4**:14.

257. Gutierrez L, Mauriat M, Guénin S, et al. **The lack of a systematic validation of reference genes: a serious pitfall undervalued in reverse transcription-polymerase chain reaction (RT-PCR) analysis in plants.** *Plant biotechnology journal* 2008, **6**:609-18.

258. Reid K, Olsson N, Schlosser J, Peng F, Lund S: **An optimized grapevine RNA isolation procedure and statistical determination of reference genes for real-time RT-PCR during berry development**. *BMC Plant Biology* 2006, **6**:27.

259. Lin YL, Lai ZX: **Reference gene selection for qPCR analysis during somatic embryogenesis in longan tree**. *Plant Science* 2010.

260. Almeida MR de, Ruedell CM, Ricachenevsky FK, et al. **Reference gene selection for quantitative reverse transcription-polymerase chain reaction normalization during in vitro adventitious rooting in *Eucalyptus globulus* Labill.** *BMC molecular biology* 2010, **11**:73.

261. Boava LP, Kuhn OJ, Pascholati SF, Di Piero RM, Luiz E: **Efeito de indutores bióticos e abióticos na atividade de quitinase e peroxidase e no controle da ferrugem causada por *Puccinia psidii* em eucalipto**. *SciELO Brasil*

262. Tusher VG, Tibshirani R, Chu G: **Significance analysis of microarrays applied to the ionizing radiation response.** *Proceedings of the National Academy of Sciences of the United States of America* 2001, **98**:5116-21.

263. Dambrauskas G, Aves SJ, Bryant JA, Francis D, Rogers HJ: **Genes encoding two essential DNA replication activation proteins, Cdc6 and Mcm3, exhibit very different patterns of expression in the tobacco BY-2 cell cycle.** *Journal of experimental botany* 2003, **54**:699-706.

264. Vandesompele J, De Preter K, Pattyn F, et al. **Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes**. *Genome Biology* 2002, **3**:1-12.

265. Andersen CL, Jensen JL, Ørntoft TF: **Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets.** *Cancer research* 2004, **64**:5245-50.

266. Allen D, Winters E, Kenna PF, Humphries P, Farrar GJ: **Reference gene selection for real-time rtPCR in human epidermal keratinocytes.** *Journal of dermatological science* 2008, **49**:217-25.

267. Coll A, Nadal A, Collado R, et al. **Natural variation explains most transcriptomic changes among maize plants of MON810 and comparable non-GM varieties subjected to two N-fertilization farming practices.** *Plant molecular biology* 2010, **73**:349-62.

268. Radonić A, Thulke S, Bae H-G, et al. **Reference gene selection for quantitative real-time PCR analysis in virus infected cells: SARS corona virus, Yellow fever virus, Human Herpesvirus-6, Camelpox virus and Cytomegalovirus infections.** *Virology journal* 2005, **2**:7.

269. Taylor S, Wakem M, Dijkman G, Alsarraj M, Nguyen M: **A practical approach to RT-qPCR-Publishing data that conform to the MIQE guidelines.** *Methods (San Diego, Calif.)* 2010, **50**:S1-5.

270. Wu SJ, Shi M, Wu JY: **Cloning and characterization of the 1-deoxy-D-xylulose 5-phosphate reductoisomerase gene for diterpenoid tanshinone biosynthesis in *Salvia miltiorrhiza* (Chinese sage) hairy roots.** *Biotechnology and applied biochemistry* 2009, **52**:89-95.

271. Carretero-Paulet L, Cunillera N, Rodrı M, et al. **Expression and Molecular Analysis of the *Arabidopsis* DXR Gene Encoding 1-Deoxy- D -Xylulose 5-Phosphate Reductoisomerase , the First Committed Enzyme of the 2-C-Methyl- D -Erythritol 4-Phosphate Pathway 1**. *Plant Physiology* 2002, **129**:1581-1591.

272. Liao Z, Chen R, Min C, et al. **A New 1-Deoxy-D-xylulose 5-Phosphate Reductoisomerase Gene Encoding the Committed-Step Enzyme in the MEP Pathway from *Rauvolfia verticillata***. *Z. Naturforsch* 2007, **62c**:296-304.

273. Yan X, Zhang L, Wang J, et al. **Molecular characterization and expression of 1-deoxy-D-xylulose 5-phosphate reductoisomerase (DXR) gene from *Salvia miltiorrhiza***. *Acta Physiologiae Plantarum* , **31**:1015-1022.

274. Yue H, Eastman PS, Wang BB, et al. **An evaluation of the performance of cDNA microarrays for detecting changes in global mRNA expression.** *Nucleic acids research* 2001, **29**:E41-1.

275. Stankovic KM, Corfas G: **Real-time quantitative RT-PCR for low-abundance transcripts in the inner ear : analysis of neurotrophic factor expression**. *Hearing Research* 2003, **185**:97-108.

276. Stahlberg A, Hakansson J, Xian X, Semb H, Kubista M: **Properties of the reverse transcription reaction in mRNA quantification**. *Clinical chemistry* 2004, **50**:509.

277. Bustin S a: **Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): trends and problems.** *Journal of molecular endocrinology* 2002, **29**:23-39.

278. Gachon C, Mingam A, Charrier B: **Real-time PCR: what relevance to plant studies?** *Journal of experimental botany* 2004, **55**:1445-54.

279. Higuchi R, Fockler C, Dollinger G: **Kinetic PCR analysis: real-time monitoring of DNA amplification reactions**. *Nature Biotechnology* 1993, **11**:1026–1030.

280. Chuaqui RF, Bonner RF, Best CJM, et al. **Post-analysis follow-up and validation of microarray experiments.** *Nature genetics* 2002, **32 Suppl**:509-14.

281. Giulietti a, Overbergh L, Valckx D, et al. **An overview of real-time quantitative PCR: applications to quantify cytokine gene expression.** *Methods (San Diego, Calif.)* 2001, **25**:386-401.

282. Czechowski T, Bari RP, Stitt M, Scheible W-R, Udvardi MK: **Real-time RT-PCR profiling of over 1400 *Arabidopsis* transcription factors: unprecedented sensitivity reveals novel root- and shoot-specific genes.** *The Plant journal : for cell and molecular biology* 2004, **38**:366-79.

283. Gutierrez L, Mauriat M, Pelloux J, Bellini C, Van Wuytswinkel O: **Towards a systematic validation of references in real-time rt-PCR.** *The Plant cell* 2008, **20**:1734-5.

284. Volkov R a, Panchuk II, Schöffl F: **Heat-stress-dependency and developmental modulation of gene expression: the potential of house-keeping genes as internal standards in mRNA expression profiling using real-time RT-PCR.** *Journal of experimental botany* 2003, **54**:2343-9.

285. Martin RC, Hollenbeck VG, Dombrowski JE: **Evaluation of Reference Genes for Quantitative RT-PCR in *Lolium perenne***. *Crop Science* 2008, **48**:1881-1887.

286. Hays D: **Development of an expressed sequence tag (EST) library for Medicago sativa**. *Plant Science* 2001, **161**:517-526.

287. Sterky F, Regan S, Karlsson J, et al. **Gene discovery in the wood-forming tissues of poplar: analysis of 5, 692 expressed sequence tags.** *Proceedings of the National Academy of Sciences of the United States of America* 1998, **95**:13330-5.

288. González-Verdejo CI, Die JV, Nadal S, et al. **Selection of housekeeping genes for normalization by real-time RT-PCR: analysis of Or-MYB1 gene expression in *Orobanche ramosa* development.** *Analytical biochemistry* 2008, **379**:176-81.

289. Hong SM, Bahn SC, Lyu A, Jung HS, Ahn JH: **Identification and Testing of Superior Reference Genes for a Starting Pool of Transcript Normalization in *Arabidopsis*.** *Plant & cell physiology* 2010, **51**:1694-1706.

290. Zhou L, Lim Q-E, Wan G, Too H-P: **Normalization with genes encoding ribosomal proteins but not GAPDH provides an accurate quantification of gene expressions in neuronal differentiation of PC12 cells.** *BMC genomics* 2010, **11**:75.

291. Kim B-R, Nam H-Y, Kim S-U, Kim S-I, Chang Y-J: **Normalization of reverse transcription quantitative-PCR with housekeeping genes in rice.** *Biotechnology letters* 2003, **25**:1869-72.

292. Løvdal T, Lillo C: **Reference gene selection for quantitative real-time PCR normalization in tomato subjected to nitrogen, cold, and light stress**. *Analytical Biochemistry* 2009, **387**:238-242.

293. McIntosh KB, Bonham-Smith PC: **The two ribosomal protein L23A genes are differentially transcribed in *Arabidopsis thaliana***. *Genome* 2005, **48**:443–454.

294. Wool IG: **Extraribosomal functions of ribosomal proteins**. *Trends in biochemical sciences* 1996, **21**:164–165.

295. Taylor MA, Arif SA, Pearce SR, et al. **Differential expression and sequence analysis of ribosomal protein genes induced in stolon tips of potato (*Solanum tuberosum* L.) during the early stages of tuberization**. *Plant physiology* 1992, **100**:1171.

296. Williams ME, Sussex IM: **Developmental regulation of ribosomal protein L16 genes in *Arabidopsis thaliana***. *The Plant Journal* 1995, **8**:65–76.

297. Revenkova E, Masson J, Koncz C, et al. **Involvement of *Arabidopsis thaliana* ribosomal protein S27 in mRNA degradation triggered by genotoxic stress.** *The EMBO journal* 1999, **18**:490-9.

298. Kim K-Y, Park S-W, Chung Y-S, et al. **Molecular cloning of low-temperature-inducible ribosomal proteins from soybean.** *Journal of experimental botany* 2004, **55**:1153-5.

299. Saez-Vasquez J, Gallois P, Delseny M: **Accumulation and nuclear targeting of BnC24, a *Brassica napus* ribosomal protein corresponding to a mRNA accumulating in response to cold treatment**. *Plant science* 2000, **156**:35–46.

300. De Angelis M, Di Cagno R, Huet C, et al. **Heat shock response in *Lactobacillus plantarum***. *Applied and environmental microbiology* 2004, **70**:1336.

301. Zhang Y, Berger S a: **Increased calcium influx and ribosomal content correlate with resistance to endoplasmic reticulum stress-induced cell death in mutant leukemia cell lines.** *The Journal of biological chemistry* 2004, **279**:6507-16.

302. Marty I, Brugidou C, Chartier Y, Meyer Y: **Growth-related gene expression in *Nicotiana tabacum* mesophyll protoplasts**. *The Plant Journal* 1993, **4**:265–278.

303. Moran DL: **Characterization of the structure and expression of a highly conserved ribosomal protein gene, L9, from pea.** *Gene* 2000, **253**:19-29.

304. Bastolla FM: **Seleção e avaliação de genes de referência para estudos de expressão gênica em *Eucalyptus***. 2007:http://www.lume.ufrgs.br/handle/10183/14358.

305. Menges M, Jager SM de, Gruissem W, Murray JAH: **Global analysis of the core cell cycle regulators of *Arabidopsis* identifies novel genes, reveals multiple and highly specific profiles of expression and provides a coherent model for plant cell cycle control.** *The Plant journal : for cell and molecular biology* 2005, **41**:546-66.

306. Joubès J, Chevalier C, Dudits D, et al. **CDK-related protein kinases in plants.** *Plant molecular biology* 2000, **43**:607-20.

307. Stone JM, Walker JC: **Plant protein kinase families and signal transduction.** *Plant physiology* 1995, **108**:451-7.

308. Umeda M, Shimotohno A, Yamaguchi M: **Control of cell division and transcription by cyclin-dependent kinase-activating kinases in plants.** *Plant & cell physiology* 2005, **46**:1437-42.

309. Reines D: **Elongation factor-dependent transcript shortening by template-engaged RNA polymerase II.** *The Journal of biological chemistry* 1992, **267**:3795-800.

310. Duchêne A-M, Giritch A, Hoffmann B, et al. **Dual targeting is the rule for organellar aminoacyl-tRNA synthetases in *Arabidopsis thaliana*.** *Proceedings of the National Academy of Sciences of the United States of America* 2005, **102**:16484-9.

311. Hoenemann C, Hohe A: **Selection of reference genes for normalization of quantitative real-time PCR in cell cultures of *Cyclamen persicum***. *Electronic Journal of Biotechnology* 2011, **14**.

312. Kamada T, Nito K, Hayashi H, et al. **Functional differentiation of peroxisomes revealed by expression profiles of peroxisomal genes in *Arabidopsis thaliana*.** *Plant & cell physiology* 2003, **44**:1275-89.

Figure Legends

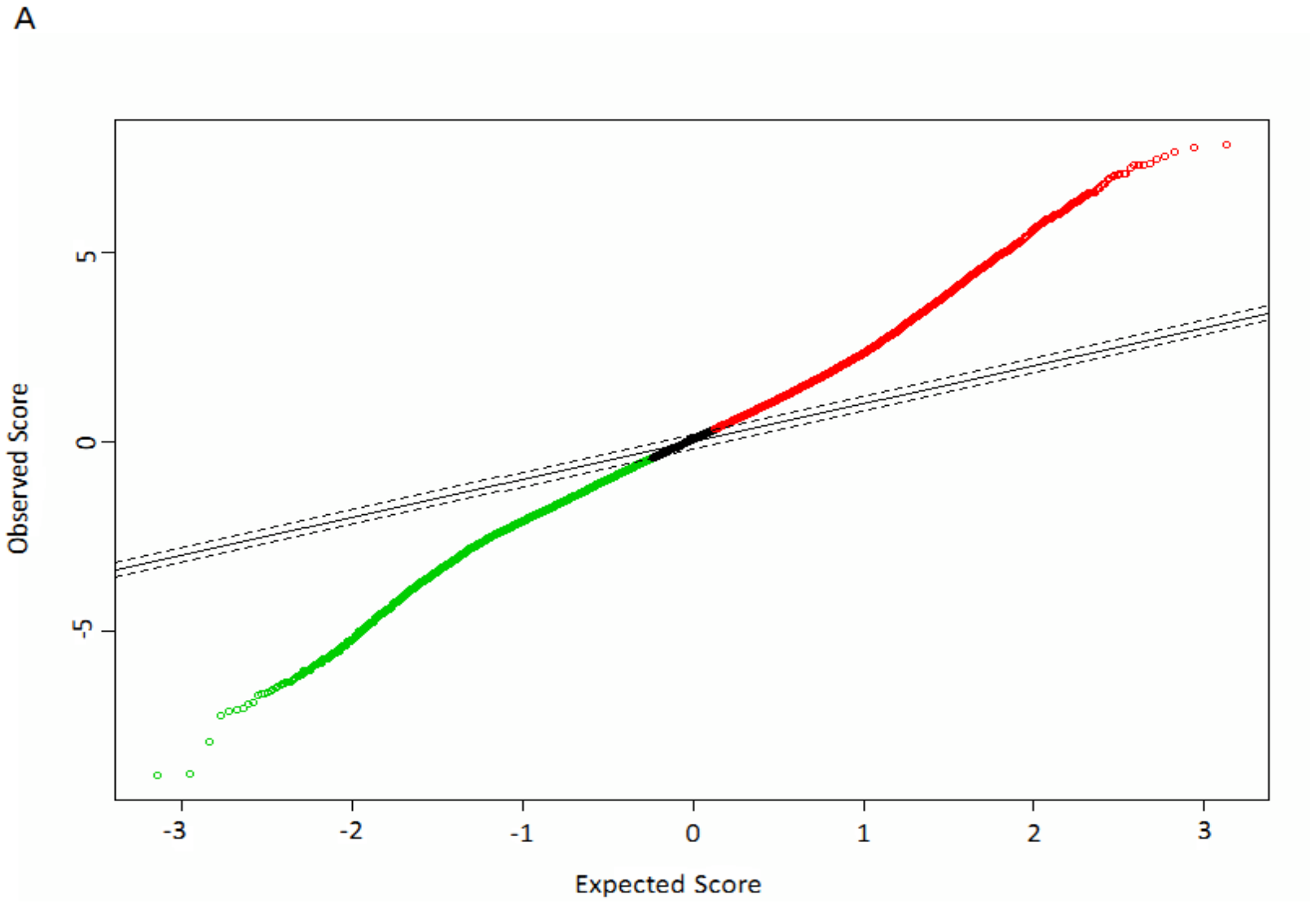
Figure 1. Expression of 21,442 *Eucalyptus* genes in *E. grandis* leaves and xylem and in *E. globulus* xylem evaluated by microarray hybridization analysis. (A) Scatter plot of the observed relative difference *d*(i) (observed score) versus the expected relative difference d_E(i) (expected score) built with the Significance Analysis of Microarray (SAM) method. The solid black line indicates the line for *d*(i) = d_E(i), where the observed relative difference is identical to the expected relative difference with a *delta* set to 0.2. Solid and dotted red and green lines represent genes whose observed relative differences are lower or higher the expected relative differences, i.e., whose expressions varied among tissues tested. (B) Three-dimensional graph generated with the Standard Deviation Microarray Analysis (SDMA) method evidencing genes expressed in positions equivalent to their overall average expression among the three conditions analyzed in the microarrays, i.e., leaves (EgrL, *z* axis) and xylem (EgrX, *x* axis) of *E. grandis* and xylem of *E. globulus* (EglX, *y* axis). (C) SDMA 3D graph representing the 50 most invariable *Eucalyptus* genes according to microarray data. Points representing selected genes tend to form a straight line since their means of expression are similar to the global average, with a standard deviation tending to zero.

Figure 2. Functional classification of of the the 50 most stable Eucalyptus genes according to the microarray data. Gene Ontology hits registered for the 50 constitutive genes that could be assigned a putative function based on Swiss-Prot query. Only known genes are shown.

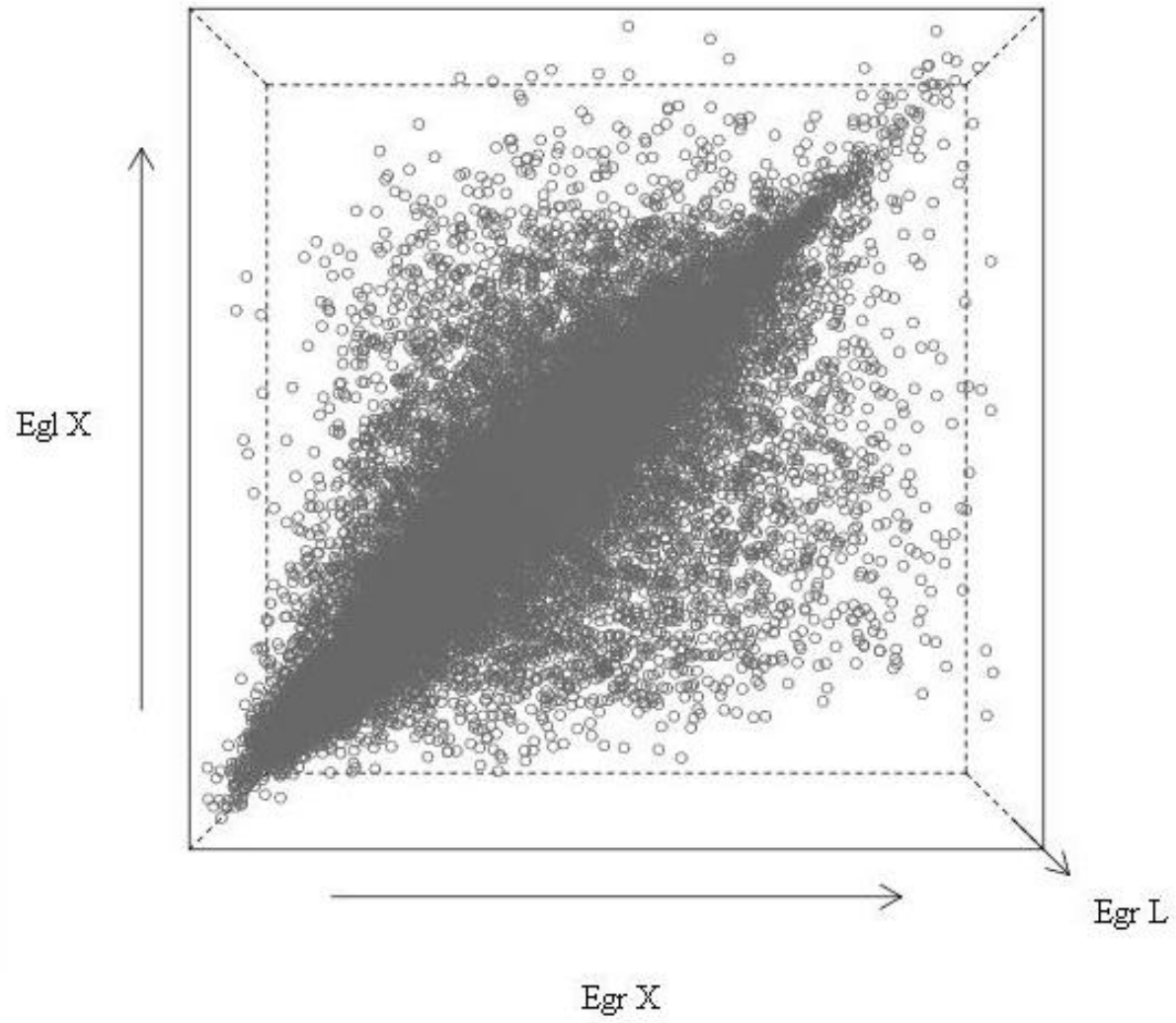
Figure 3. Expression stability of candidate genes evaluated by RT-qPCR in all tissues/organs and species of *Eucalyptus* analyzed. Results were analyzed according to algorithms *geNorm* and *NormFinder* and represent the general average values of expression. Lower the values, more stable the gene expression. (A) Average *M* stability values of candidate gene expressions calculated with the *geNorm* algorithm. (B) Pairwise variation (*V*) values calculated with the *geNorm* algorithm in order to estimate the optimal number of reference genes necessary for accurately normalize the expressions of genes of interest. Values lower than 0.15 indicate that no additional genes are required for the normalization of expression in organs/tissues studied. (C) Stability values of gene expressions calculated with the *NormFinder* algorithm.

Figure 4. Relative expression of the isoprenoid biosynthetic gene *dxr* in different tissues/organs of *Eucalyptus* by RT-qPCR and normalization with different reference gene pairs. Gene pairs employed as references are indicated in the top of the graphics. Average values of reference gene relative expressions in the different tissues were set to one (1) in order to normalize the *dxr* expressions. (A) Expression patterns of the *dxr* gene in leaves, flowers and xylem tissues of *E. grandis*. (B) Expression patterns of the *dxr* gene in xylem tissues of *E. grandis*, *E. globulus* and *E. pellita.*

Figure 1



B



C

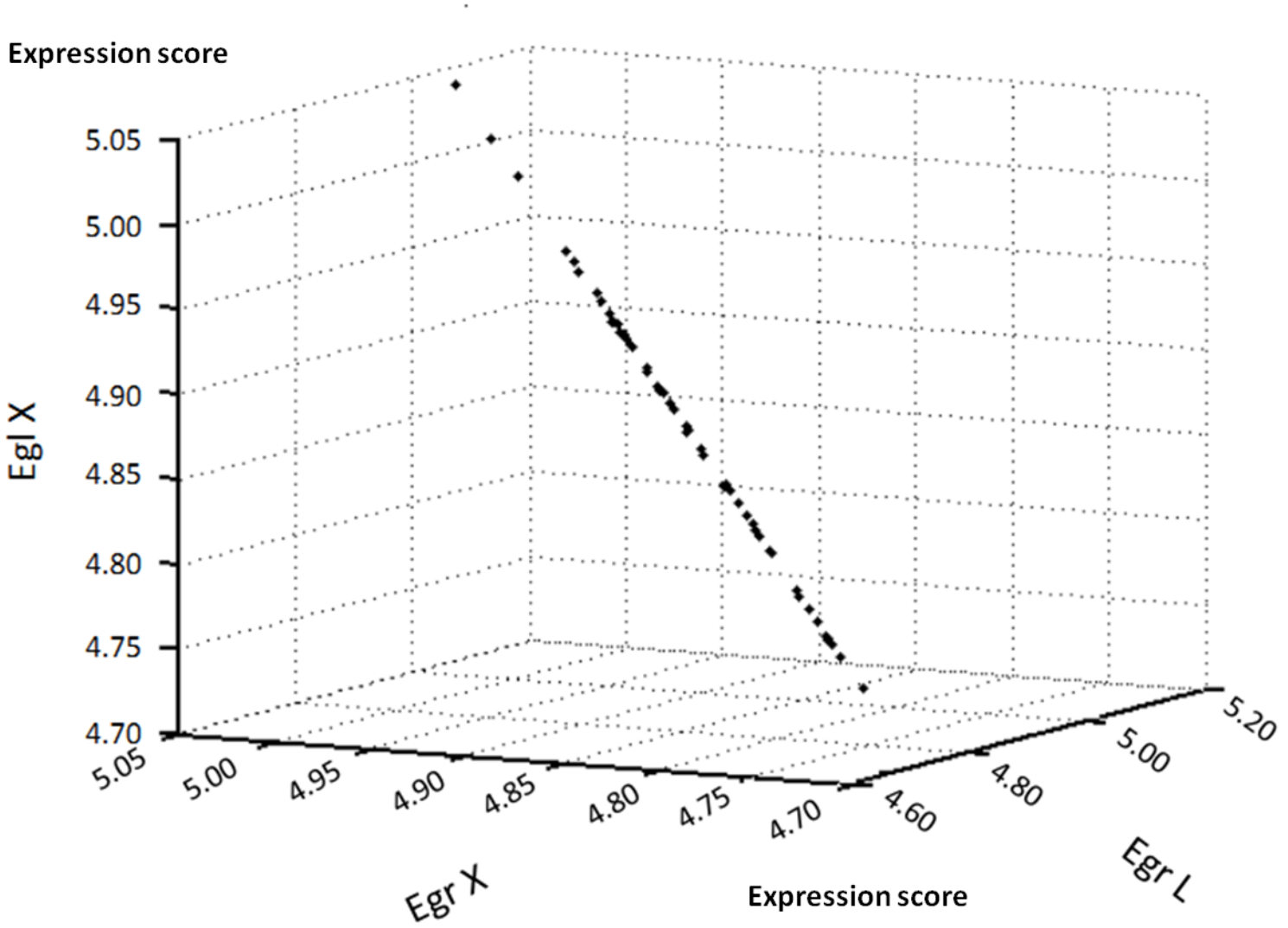

Expression score
5.05
5.00
4.95
4.90
4.85
4.80
4.75
4.70
Egl X
5.05
5.00
4.95
4.90
4.85
4.80
4.75
4.70
Egr X
4.60
4.80
5.00
5.20
Egr L
Expression score

Figure 2

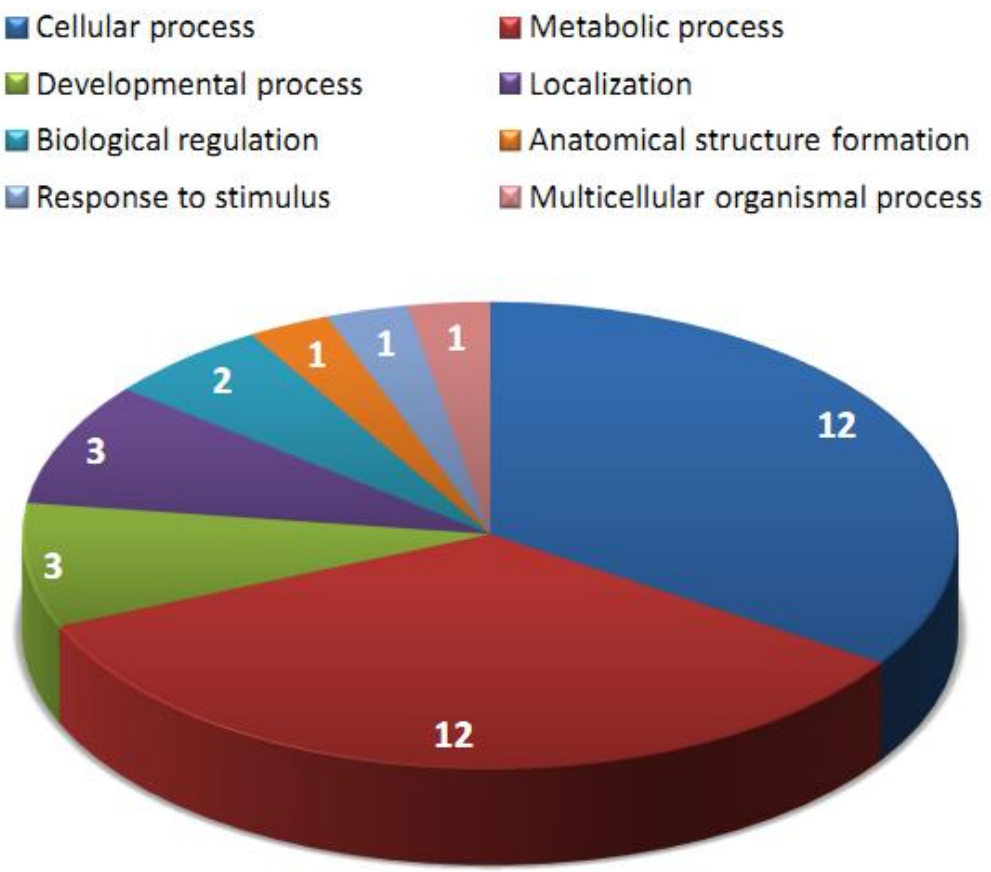
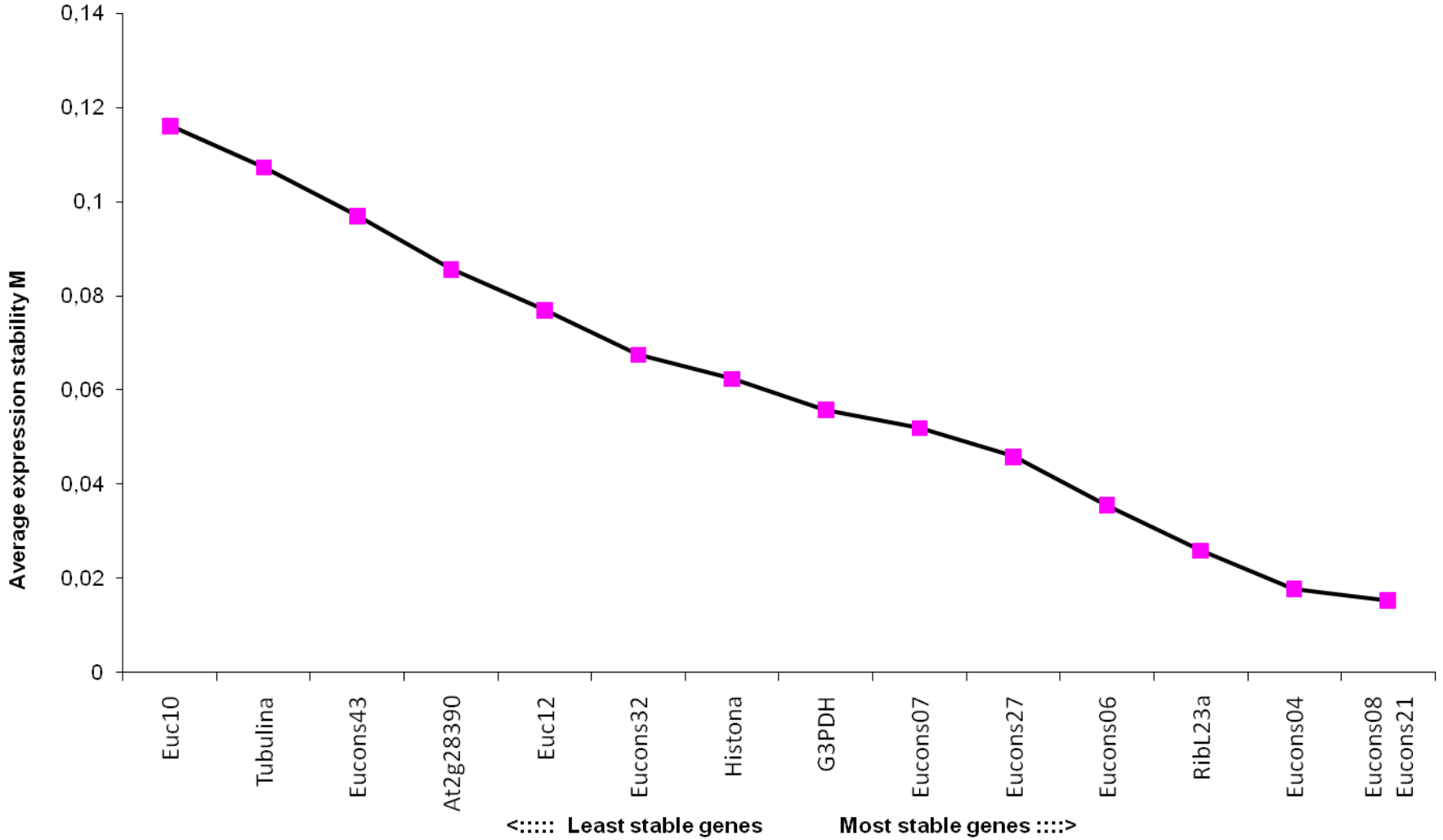
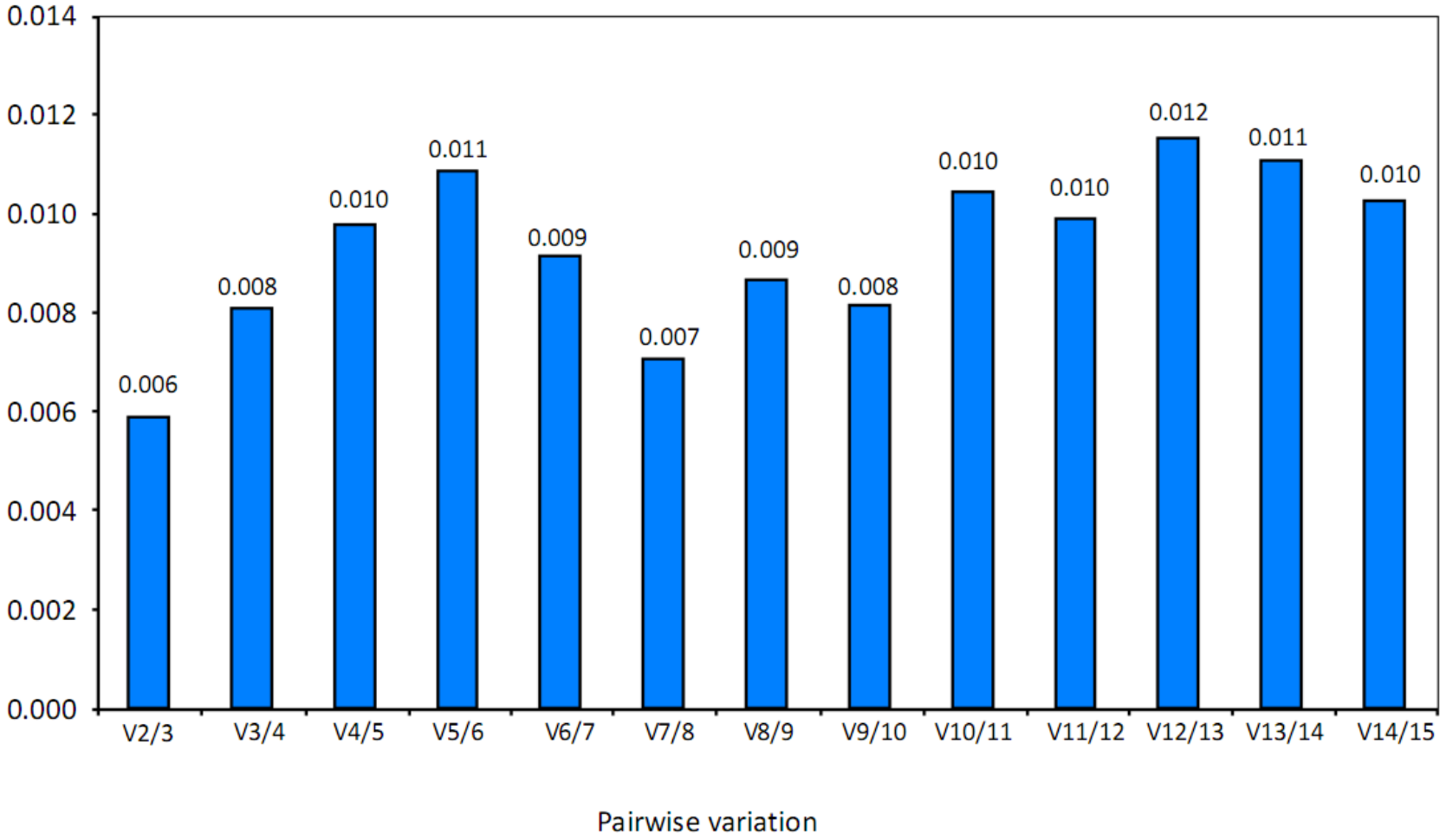

Cellular process
Metabolic process
Developmental process
Localization
Biological regulation
Anatomical structure formation
Response to stimulus
Multicellular organismal process
12
12
3
3
2
1
1
1

Figure 3 **A**


0,14
0,12
0,1
0,08
0,06
0,04
0,02
0
Average expression stability M
Euc10
Tubulina
Eucons43
At2g28390
Euc12
Eucons32
Histona
G3PDH
Eucons07
Eucons27
Eucons06
RibL23a
Eucons04
Eucons08
Eucons21
<::::: Least stable genes
Most stable genes ::::>

B


0.014
0.012
0.010
0.008
0.006
0.004
0.002
0.000
0.006
0.008
0.010
0.011
0.009
0.007
0.009
0.008
0.010
0.010
0.012
0.011
0.010
V2/3
V3/4
V4/5
V5/6
V6/7
V7/8
V8/9
V9/10
V10/11
V11/12
V12/13
V13/14
V14/15
Pairwise variation

C

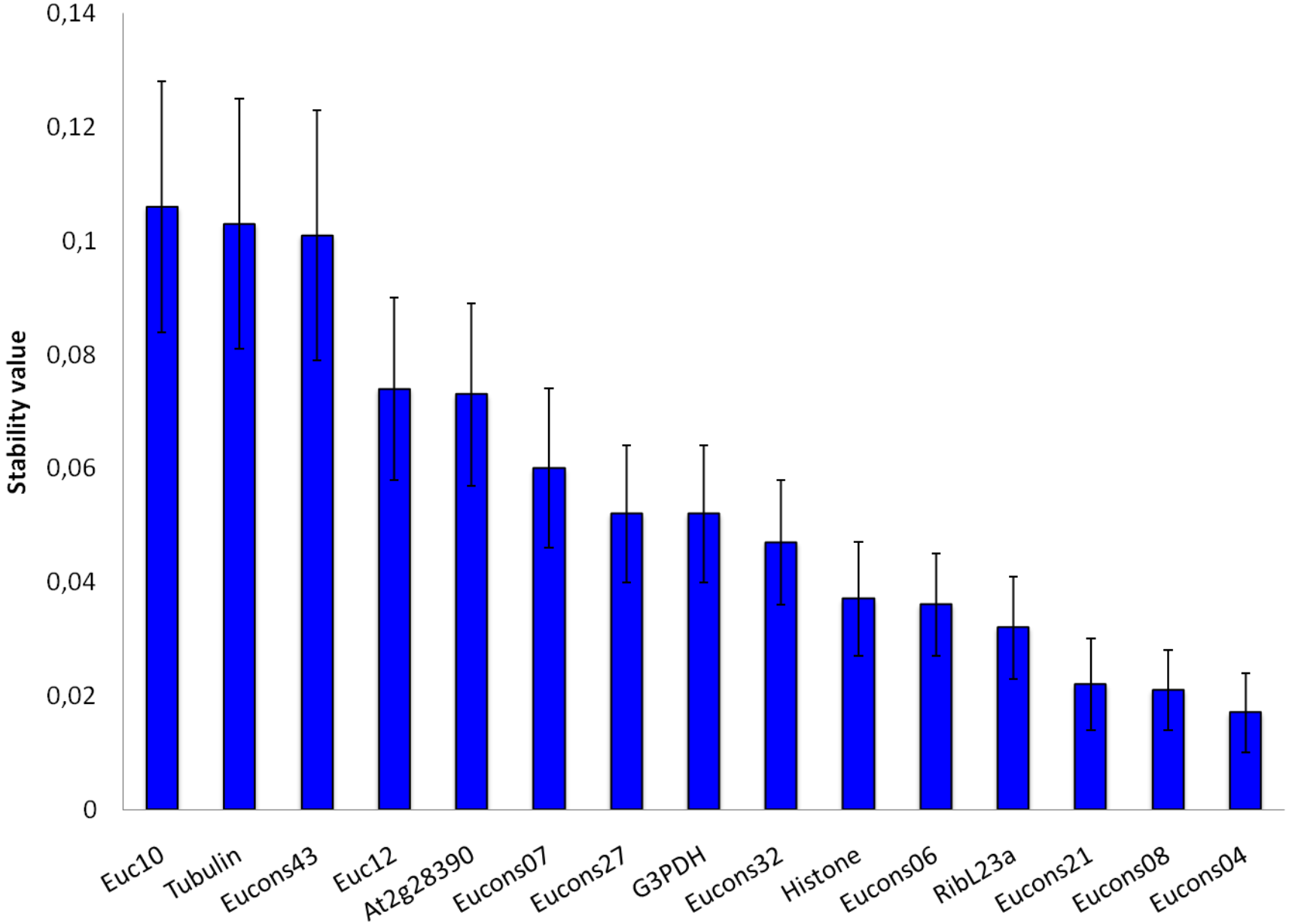
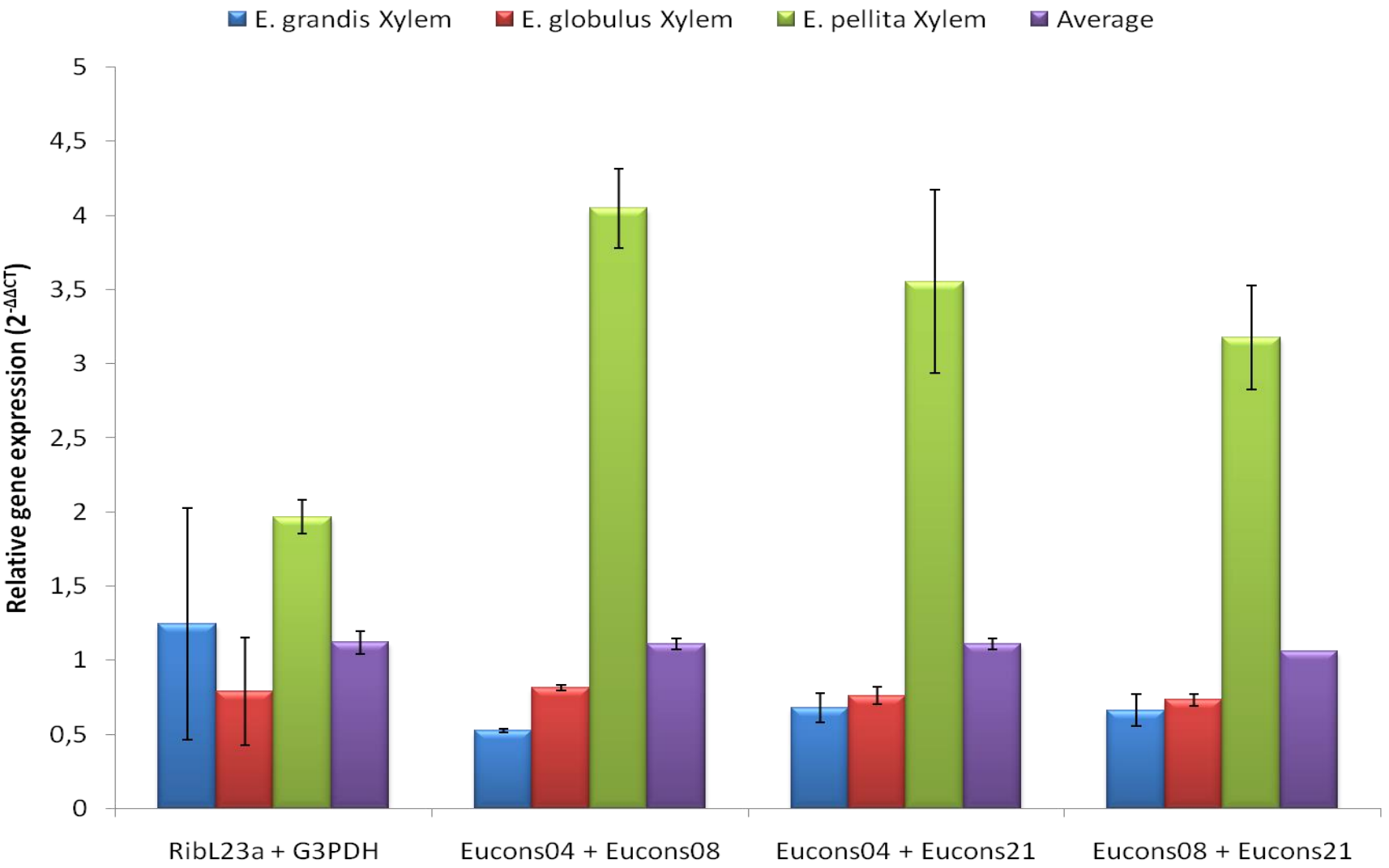
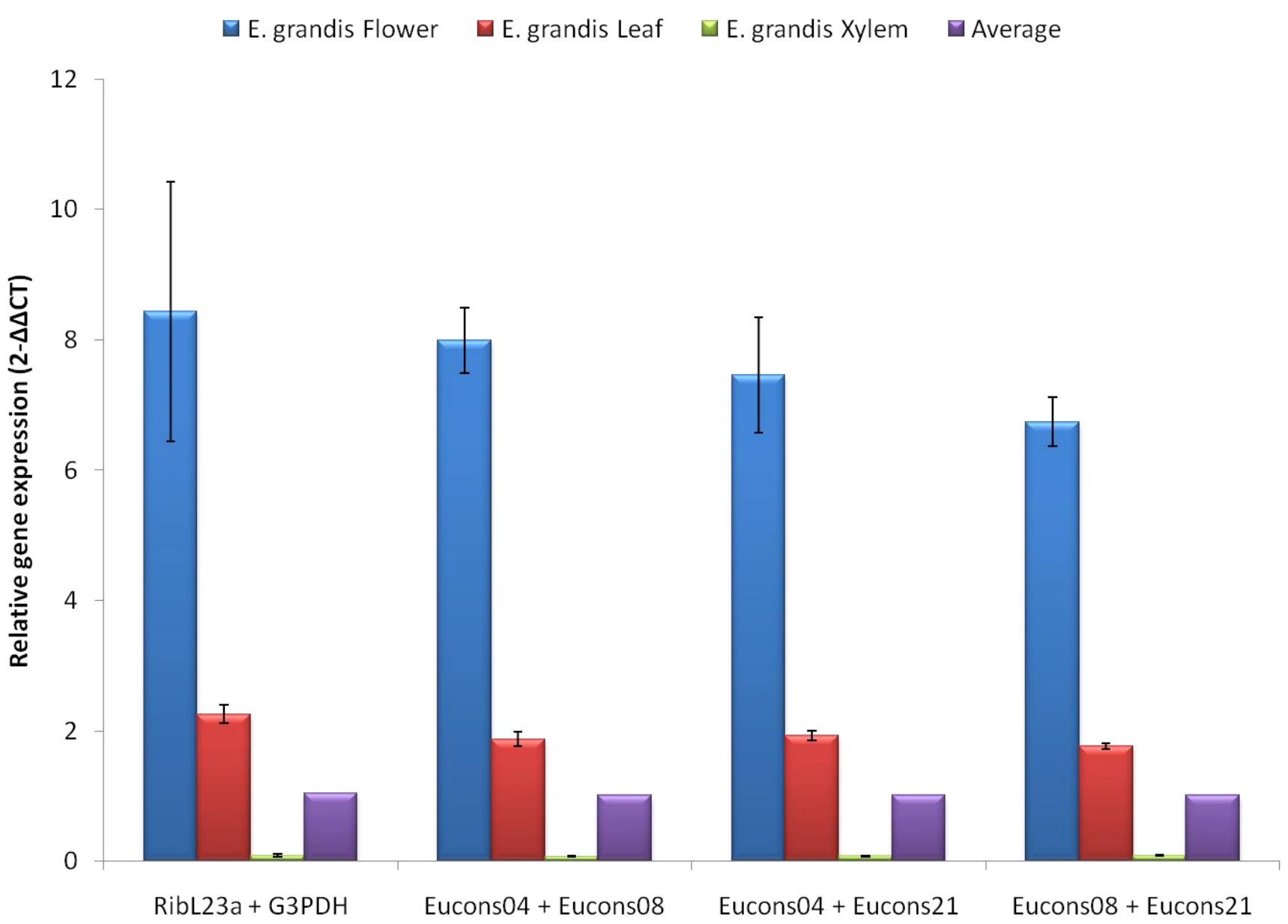

0,14
0,12
0,1
0,08
0,06
0,04
0,02
0
Stability value
Euc10
Tubulin
Eucons43
Euc12
At2g28390
Eucons07
Eucons27
G3PDH
Eucons32
Histone
Eucons06
RibL23a
Eucons21
Eucons08
Eucons04

Figure 4 **A**


E. grandis Xylem
E. globulus Xylem
E. pellita Xylem
Average
Relative gene expression (2-ΔΔCT)
5
4,5
4
3,5
3
2,5
2
1,5
1
0,5
0
RibL23a + G3PDH
Eucons04 + Eucons08
Eucons04 + Eucons21
Eucons08 + Eucons21

B


E. grandis Flower
E. grandis Leaf
E. grandis Xylem
Average
Relative gene expression (2-ΔΔCT)
12
10
8
6
4
2
0
RibL23a + G3PDH
Eucons04 + Eucons08
Eucons04 + Eucons21
Eucons08 + Eucons21

Tables

Table 1. Fifty most stable *Eucalyptus* genes selected from microarray data analysis employing the SDMA and SAM statistical algorithms. GenBank accession codes and putative functional identity based on BLAST analysis are indicated along with the estimated value (E-value). Results of the statistical analysis performed are indicated: standard deviations (SD) for the SDMA method, and fold change and final score of the SAM method. Genes were ranked from highest to lowest stability for both methods. Lines shaded in gray represent genes selected for validation via RT-qPCR analysis.

Table 2. Primer sequences (5'-3') employed in the RT-qPCR analysis of candidate reference genes. Forward and reverse primer sequences are provided. Estimated amplicon lengths are indicated.

Table 3. Genes traditionally employed as references in RT-qPCR analysis. Gene abbreviations, GenBank accession codes and functional identity are indicated. GenBank accessions to *Eucalyptus* best hit sequence are also presented. Based on the *Eucalyptus* sequences, primers were designed as shown along with the expected amplicon lengths.

Table 4. *Eucalyptus* reference genes for normalization and their expression stability values calculated by the *NormFinder* software.

Table 1

Gene	Gene ID	BLAST Annotation	SDMA		SAM	
			Standard Deviation	Ranking	Fold change	Score (d)	Ranking
Eucons01	emb|CAY47298.1|	serine transporter [*Pseudomonas fluorescens SBW25]*	0.000082	1	0,999995685	0,00661473	48
Eucons02		No hit	0.00029	2	1,000005363	0,007716502	23
Eucons03	gb|EEY18801.1|	DNA damage checkpoint protein rad24 [*Verticillium albo-atrum VaMs.102*]	0.000365	3	0,999993841	0,008958654	41
Eucons04	gb|EEF43392.1|	cdk8, putative [*R. communis*]	0.000461	4	1,000006755	0,009261212	21
Eucons05		No hit	0.000509	5	1,000008973	0,010259132	46
Eucons06	gb|EEF44719.1|	Plastidic ATP/ADP-transporter, putative [*R. communis*]	0.000549	6	1,000009876	0,012322456	40
Eucons07	gb|EEF03117.1|	ABC transporter family protein [*P. trichocarpa*]	0.000561	7	1,000010164	0,01241657	8
Eucons08	gb|EEF33688.1|	transcription elongation factor s-II, putative [*R. communis*]	0.000582	8	1,000013416	0,012556064	15
Eucons09	gb|EEF42371.1|	nucleic acid binding protein, putative [*R. communis*]	0.000665	9	1,000015637	0,013553234	9
Eucons10	gb|ACG37397.1|	anther-specific proline-rich protein APG [*Z. mays*]	0.000678	10	1,000018989	0,015574381	2

Gene	Gene ID	BLAST Annotation	SDMA		SAM		
			Standard Deviation	Ranking	Fold change	Score (d)	Ranking
Eucons11		No Hit	0.000694	11	1,0000296	0,016330796	44
Eucons12		No hit	0.000694	12	0,999973005	0,016437714	7
Eucons13		No hit	0.000732	13	0,999966	0,016807148	17
Eucons14	gb|AAM52237.1|	senescence/dehydration-associated protein-related [*A.thaliana*]	0.000734	14	1,000035875	0,017118153	36
Eucons15	emb|CAA65477.1|	lipid transfer protein [*Prunus dulcis*]	0.000785	15	1,000039015	0,018024764	5
Eucons16	gb|EEF44560.1|	F-box and wd40 domain protein, putative [*R. communis*]	0.000791	16	0,999960301	0,01818155	35
Eucons17		No hit	0.000794	17	1,000045767	0,020617825	34
Eucons18		No hit	0.000800	18	0,999960851	0,021328632	4
Eucons19		No hit	0.000821	19	1,000046587	0,021445011	1
Eucons20	gb|EEE97842.1|	chromatin remodeling complex subunit [*P. trichocarpa*]	0.000879	20	1,000054302	0,021978112	14

Gene	Gene ID	BLAST Annotation	SDMA		SAM		
			Standard Deviation	Ranking	Fold change	Score (d)	Ranking
Eucons21	gb|EEF48129.1|	aspartyl-tRNA synthetase, putative [*R. communis*]	0.000913	21	0,99994759	0,022500691	6
Eucons22		No hit	0.000931	22	1,000052414	0,02331048	28
Eucons23	dbj|BAB02414.1|	chloroplast nucleoid DNA binding protein-like [*A. thaliana*]	0.000970	23	1,000055076	0,02331048	10
Eucons24	gb|EEF45372.1|	conserved hypothetical protein [*R. communis*]	0.000992	24	0,999948713	0,024728932	19
Eucons25		No hit	0.001028	25	0,999941348	0,024874888	11
Eucons26	gb|EEF45384.1|	vacuole membrane protein, putative [*R. communis*]	0.001030	26	1,000068819	0,024983297	16
Eucons27	gb|EEF44734.1|	peroxisome biogenesis factor, putative [*R. communis*]	0.001063	27	1,000059599	0,025183557	13
Eucons28		No hit	0.001085	28	0,999931481	0,025250589	12
Eucons29	gb|EER99842.1|	hypothetical protein SORBIDRAFT_02g041780 [*S.bicolor*]	0.001158	29	0,999914659	0,025278616	20
Eucons30	gb|EEF45541.1|	sentrin/sumo-specific protease, putative [*R.communis*]	0.001175	30	1,000062543	0,025638209	50

Gene	Gene ID	BLAST Annotation	SDMA		SAM		
			Standard Deviation	Ranking	Fold change	Score (d)	Ranking
Eucons31	emb|CAP42856.1|	SsrA-binding protein [*Bordetella petrii*]	0.001175	31	0,999931592	0,02654049	18
Eucons32	gb|EDS90429.1|	nitrogen regulation protein NR(II) [*Escherichia albertii TW07627*]	0.001181	32	1,000079957	0,026547459	26
Eucons33		No hit	0.001189	33	1,000081012	0,026669523	30
Eucons34	gb|EEE90904.1|	predicted protein [*P. trichocarpa*]	0.001256	34	1,000078558	0,026841264	37
Eucons35		No hit	0.001257	35	1,000079192	0,027114749	38
Eucons36	ref|NP_565080.1|	mitochondrial transcription termination factor-related / mTERF-related [*A. thaliana*]	0.001263	36	0,999911838	0,027388068	24
Eucons37		No hit	0.001273	37	0,999915438	0,027861819	45
Eucons38	emb|CAN64407.1|	hypothetical protein [*V. vinifera*]	0.001276	38	0,999891043	0,027872741	25
Eucons39		No hit	0.001286	39	0,99990712	0,028070603	33
Eucons40		No hit	0.001293	40	0,999908155	0,028195758	32

Gene	Gene ID	BLAST Annotation	SDMA		SAM		
			Standard Deviation	Ranking	Fold change	Score (d)	Ranking
Eucons41		No hit	0.001303	41	1,000098979	0,028550233	29
Eucons42		No hit	0.001361	42	0,999893249	0,028605047	27
Eucons43	gb|EEF46905.1|	Serine/threonine-protein kinase PBS1, putative [*R. communis*]	0.001365	43	1,000105741	0,028682612	43
Eucons44	gb|EEF48108.1|	pollen specific protein s f21, putative [*R. communis*]	0.001378	44	1,000117513	0,030661159	42
Eucons45		No hit	0.001385	45	0,9999055	0,030782861	49
Eucons46	gb|ACM45716.1|	class IV chitinase [*Pyrus pyrifolia*]	0.001428	46	1,000104153	0,031264117	22
Eucons47	gb|EEE86166.1|	f-box family protein [*P. trichocarpa*]	0.001432	47	1,00010279	0,031514942	39
Eucons48		No hit	0.001613	48	1,00012433	0,031516084	47
Eucons49	gb|EEF03628.1|	predicted protein [*P. trichocarpa*]	0.001848	49	0,999886858	0,031626533	31
Eucons50		No hit	0.001303	50	1,000098979	0,031987517	29

Table 2

Gene	Forward primer	Reverse primer	Amplicon size (pb)
Eucons01	TGGTGCTGACGGTGATGTTCTTCT	AAGGATTTGGTGATCGCCACCAGT	178
Eucons04	TACAAGCGCTGTTGATATGTGGGC	TTGCCAATGAGGCGGATTCACAAG	196
Eucons06	TCCTCTGTCCACAAATGGGTTCCA	TCACCAAAGACAGGCTGACCATCA	141
Eucons07	AAGCCTCATTGGCTGGCTCACATA	TCAGCACAAGAGCTCCACCATCAT	153
Eucons08	TCCAATCCGAGTCGCTGTCATTGT	TGATGAGCCTCTCTGGTTTGACCT	152
Eucons15	AAGTGAGAGCAAAGATGGAGCGCA	GACCATATTACACGACGCATCGCA	154
Eucons21	AGAGGTGAAATTCCAGAAGCCCGT	CTTCCCTTTGGCTTCCGCCAATTA	155
Eucons27	CATTTCATGCTGCTGTTGGCCGTT	AGTCCACCAACATCATCCCATCCA	184
Eucons32	GACAACGTGCGGTTGATTCGTGAT	ACGCAGAATGATTTCACCGCCTTC	144
Eucons43	TATTTCTCCTGTTTCGCTCCGGGT	TACCATCTCTTTGTGCTCTGCGCT	166

Table 3

Gene	Gene ID	Annotation	Genolyptus Code	Primers	Amplicon (pb)	Reference
At2g28390	AT2G28390.1	SAND family protein [*A. thaliana*]	CL10532Contig1	CCATTCAACACTCTCCGACA TGTGTGACCCAGCAGAGTAAT	143	[245]
G3PDH	AT1G13440.1	glyceraldehyde-3-phosphate dehydrogenase [*A. thaliana*]	CL1Contig558	TTGGCATTGTTGAGGGTCTA AAGCAGCTCTTCCACCTCTC	107	[263]
Histone	AT5G02570.1	histone H2B, putative [*A. thaliana*]	CL1Contig1328	GAGCGTGGAGACGTACAAGA GGCGAGTTTCTCGAAGATGT	127	[207]
RibL23a	AT2G39460.2	Ribosomal protein L23A [*A. thaliana*]	CL501Contig2	AAGGACCCTGAAGAAGGACA CCTCAATCTTCTTCATCGCA	128	[207]
Tubulin	AT1G50010.1	TUA2; structural constituent of cytoskeleton [*A. thaliana*]	CL1Contig134	GCAAGTACATGGCTTGCTGT CACACTTGAATCCTGTTGGG	132	[207]
Euc10	AT3G07640.1	unknown protein [*A. thaliana*]	CL4795Contig1	AGGAGTCCTTCGAGCTTCC CAGCACGGACACCTGATAAA	110	[304]
Euc12	AT1G32790.1	RNA-binding protein, putative [*A. thaliana*]	CL3180Contig1	GCGTGGTTCTTGGATCACTA TGGTGACAAAGTCAGGTGCT	114	[260, 304]

Table 4

1 Group						2 Groups	
Leaves		Xylem		All organs		Leaves + Xylem	
Ranking	SV+ SD	Ranking	SV+ SD	Ranking	SV+ SD	Ranking	SV
Eucons04	0.008 ± 0.010	*Eucosn27*	0.010 ± 0.010	*Eucons04*	0.017 ± 0.007	*Eucons04*	0.011
Eucons08	0.018 ± 0.009	*Eucons07*	0.017 ± 0.010	*Eucons08*	0.021 ± 0.007	*Eucons08*	0.016
Eucons32	0.023 ± 0.010	*Eucons06*	0.023 ± 0.010	*Eucons21*	0.022 ± 0.008	*Eucons21*	0.019
Eucons21	0.027 ± 0.011	Histone	0.024 ± 0.011	RibL23a	0.032 ± 0.009	RibL23a	0.020
RibL23a	0.033 ± 0.012	*Eucons21*	0.025 ± 0.011	*Eucons06*	0.036 ± 0.009	*Eucons06*	0.036
G3PDH	0.035 ± 0.013	RibL23a	0.027 ± 0.011	Histone	0.037 ± 0.010	Histone	0.037
Eucons06	0.038 ± 0.014	*Eucons04*	0.028 ± 0.012	*Eucons32*	0.047 ± 0.011	*Eucosn27*	0.045
Histone	0.045 ± 0.016	*Eucons08*	0.029 ± 0.012	G3PDH	0.052 ± 0.012	*Eucons32*	0.046
Eucons07	0.051 ± 0.017	*Euc12*	0.036 ± 0.013	*Eucosn27*	0.052 ± 0.012	G3PDH	0.048
At2g28390	0.054 ± 0.018	*Eucons32*	0.050 ± 0.017	*Eucons07*	0.06 ± 0.014	*Eucons07*	0.058
Eucosn27	0.059 ± 0.020	G3PDH	0.056 ± 0.019	At2g28390	0.073 ± 0.016	*Euc10*	0.062
Tubulin	0.059 ± 0.020	At2g28390	0.058 ± 0.020	*Euc12*	0.074 ± 0.016	At2g28390	0.067
Euc12	0.069 ± 0.023	Tubulin	0.071 ± 0.023	*Eucons43*	0.101 ± 0.022	*Euc12*	0.068
Euc10	0.074 ± 0.024	*Eucons43*	0.088 ± 0.029	Tubulin	0.103 ± 0.022	*Eucons43*	0.083
Eucons43	0.086 ± 0.028	*Euc10*	0.138 ± 0.044	*Euc10*	0.106 ± 0.023	Tubulin	0.094
Best combination of 2 genes						*Eucons04* *Eucons08*	0.009

CONCLUSÕES GERAIS

- Um total de 3.190 ESTs foram geradas ao todo, totalizando 1.641 ESTs, distribuídas em 120 *clusters* e 2.830 *singletons*;
- As sequências foram assim distribuídas entre as bibliotecas: 69 unigenes estão presentes na biblioteca de sépalas/pétalas, 136 em anteras/estames, 68 em pistilos e 130 em carpelos/receptáculo floral;
- O maior número de ESTs identificadas referem-se a genes de expressão constitutiva ou a genes potencialmente relacionados aos sinais fisiológicos de senescência da flor;
- 22 sequencias foram identificadas, pela presença do domínio conservado Dof, como pertencentes a família de transcrição Dof;
- Tipicamente, a arquitetura dos genes *Dof* exibiram somente um ou dois exons em sua estrutura gênica, totalizando 15 sequencias. As demais sequencias apresentaram de três a seis exons, demosntrando a alta variabilidade dos genes *Dof* de *E. grandis*;
- Foi observado um padrão bastante diverso na expressão dos genes *Dof* de *E. grandis* avaliados por RT-qPCR, sugerindo seu envolvimento em diferentes vias metabólicas na planta;
- 50 genes potencialmente constitutivos foram selecionados por análises *in silico* dos dados gerados por microarranjos de DNA de folhas e xilema de *E. grandis* e xilema de *e. globulus*;
- 15 destes genes foram selecionados para análises de RT-qPCR em diferentes órgãos (flor, folha e xilema) de seis species de *Eucalyptus*;
- De acordo com as análises de estabilidade de expressão realizada no programa NormFinder e GeNorm, o par de genes Eucons04 e Eucons08 são os melhores genes refência para análises em folhas e xilema, ou em folhas isoladamente. Para análises em tecido xilemático, os genes Eucons07 e Eucons27 se mostraram os mais estáveis. Considerando todos os tecidos e espécies de *Eucalyptus* analisadas, os genes *Eucons04, Eucons08* e *Eucons21* são os genes que apresentaram maior estabilidade na expressão e consequentemente, são

os melhores genes para avaliação da expressão em tecidos e espécies de *Eucalyptus.*

References

1. Pereira JCD, Sturion JA, Higa AR, Higa RCV, Shimizu JY: *Caracteristicas da madeira de algumas espécies de eucalipto plantadas no Brasil*. Colombo PR: Embrapa; 2000:3-114.

2. Queiroz LR de S, Barrichelo LEG: *Eucalipto um século no Brasil*. São Paulo - Brasil: 2007.

3. Mora AL, Garcia CH: **A Cultura do Eucalipto no Brasil**. 2000.

4. Eldridge K, Davidson J, Harwood C, Van Wyk G: *Eucalypt domestication and breeding*. Oxford Univ Pr on Demand; 1994:288.

5. Santos SN: **Genes de lignificação em *Eucalyptus*: estrutura e diversidade genética dos genes 4cl e ccoaomt.** 2005.

6. Oliveira LA: **Análise transcricional dos genes ISA1 , NFS1 e ISU1 de *Eucalyptus grandis* sob estresse**. *Stress: The International Journal on the Biology of Stress* 2008:1-88.

7. Auer CG: **A ocorrência do cancro do Eucalipto nos Estados do Paraná e de Santa Catarina**. *Boletim de Pesquisa Florestal* 1996:81-83.

8. Siqueira L: **Variabilidade genética de população selecionada e teste de paternidade de uma progênie de polinização aberta de *Eucalyptus grandis* Hill ex Maiden**. *Agricultural and Forest Entomology* 2009, **74**:40.

9. Devide AC: **Avaliação da adaptabilidade e da estabilidade fenotípica de progenies de *Eucalyptus pellita* F . Muell , introduzidas da Austrália**. 1992.

10. Sousa VA, Higa RCV: **Fenologia de florescimento e frutificação de *Eucalyptus dunnii*.** 1991:9-19.

11. Scanavaca Junior L: **Caracterização silvicultural, botânica e tecnológica do *Eucalyptus urophylla* S.T. Blake e de seu potencial para utilização em serraria**. 2001.

12. FERREIRA M: **Melhoramento ea silvicultura intensiva clonal**. *Scientia Florestalis* 1992, **45**:22–30.

13. Schneider PR, Coelho MCB, Zanon ML, Finger CAG, Klein JEM: **Equações de volume para *E. dunnii* Maiden, determinadas para depressão central do estado do Paraná e Rio Grande do Sul**. *Ciência Rural* 1997, **27**:425-428.

14. Xavier AA, Sanfuentes EV, Jughans DT, Alfenas AC: **Resistência de *Eucalyptus globulus* E *Eucalyptus nitens* à ferrugem (*Puccinia psidii*)**. *Revista Árvore* 2007, **31**:731-735.

15. Pezzutti RV, Scnumacher MV, Hoppe JM: **Crescimento de mudas de *Eucalyptus globulus* em resposta à fertilização NPK**. *Ciência Florestal* 1997, **9**:117-125.

16. Rocha ME do N, Santos CL dos: **O uso comercial e popular do eucalipto *Eucalyptus globulus* Labill - Myrtaceae**. *Saúde e Ambiente* 2007, **2**:23-24.

17. Boland DJ, Brooker MIH, Chippendale GM, et al. *Forest trees of Australia*. Csiro; 2006:715.

18. Dibax R: **Transformação genética de *Eucalyptus saligna* com o gene P5CSF129A via *Agrobacterium tumefaciens***. 2007.

19. Ferreira M: **Escolha de espécies de eucalipto**. *Circular Técnica IPEF* 1979, **47**:1–30.

20. Andrade EN: *O Eucalipto*. Rio Claro SP: 1961:667.

21. Carvalho ADF, Geraldi IO: **Histórico do melhoramento genético de eucalipto no Brasil**. In 2006:11-13.

22. Xavier A, Luiz R: **Evolução da silvicultura clonal de *Eucalyptus* no Brasil**. 2010, **34**:93-98.

23. Giri CC, Shyamkumar B, Anjaneyulu C: **Progress in tissue culture, genetic transformation and applications of biotechnology to trees: an overview**. *Trees - Structure and Function* 2004, **18**:115-135.

24. PASQUALI G, ZANETTINI M, BORÉM A: **Transgênese florestal**. In *BORÉM, A. Biotecnologia florestal. Viçosa: UFV*. Viçosa - MG: 2007:317–334.

25. Sartoretto LM, Saldanha CW, Corder MPM: **Melhoramento genético de espécies florestais**. *Revista Madeira* 2011, **125**:1-4.

26. Grattapaglia D: **Integrating genomics into *Eucalyptus* breeding.** *Genetics and molecular research : GMR* 2004, **3**:369-79.

27. Carrer H: **Seeing the FORESTs for the trees**. *Genetics and Molecular Biology* 2005, **28**:2004-2005.

28. Myburg Z, Grattapaglia D, Potts B, et al. **Sequencing of the *Eucalyptus* Genome: a poposal to DOE-JGI Prepared by**. *Genomics* 2006.

29. ABRAF: **Anuário estatístico da ABRAF 2010 ano base 2009**. *ABRAF - Brasília* 2010, **54**:140p.

30. Bracelpa: **Estatísticas Bracelpa - Relatório Anual 2009/2010**. 2010, **4124097764**:60.

31. Soares NS, Rezende AM, Silva ML, Cordeiro SA, Moura AD de: **Mercado de produtos florestais se aquece e perspectivas novas surgem para produtos certificados**. 2010:1-7.

32. Sociedade Brasileira de Silvicultura: **Fatos e Números do Brasil Florestal**. 2008.

33. Lorenzi H: *Árvores exóticas no Brasil: madeireiras, ornamentais e aromáticas*. Instituto Plantarum de Estudos da Flora; 2003.

34. Richards A: *Plant breeding systems*. Garland Science; 1986:529.

35. Boland DJ, Brooker MIH, Chippendale GM, et al. *Forest trees of Australia*. Fifth edit. Melbourne: Csiro Publishing; 1984:687.

36. Paludzyszyn Filho E, Pacheco AR, Dittmar H, Cordeiro CA: *Estratégias para o Melhoramento de Eucaliptos Trocipais na Embrapa*. Colombo - PR: 2004, **Documentos**.

37. Aguiar IB de, Kageyama PY: **Desenvolvimento floral de *Eucalyptus grandis* HILL EX Maiden em Mogi Guaçu - SP**. *IPEF* 1987, **37**:5-11.

38. Williams J, Woinarski J: *Eucalypt ecology individuals to eosystems*. Cambridge University Press; 1997, **11**:413.

39. Assis FT: **Melhoramento genético do eucalipto**. *Informe Agropecuário* 1986, **12**:36-46.

40. Hodgson LM: **Some aspects o flowering and reproductive behaviour in *Eucalyptus grandis* (Hill) Maiden at J.D.M. Keet Forest Reearch station**. *South African Forestry Journal* 1976, **98**:32-43.

41. Schumacher MV, Calil FN, Vogel HLMV: *Silvicultura Aplicada*. Santa Maria - RS: 2005:120.

42. Fobel M, Lynch DV, Thompson JE: **Membrane Deterioration in Senescing Carnation Flowers ' COORDINATED EFFECTS OF PHOSPHOLIPID DEGRADATION AND THE ACTION OF**. *Photosynthetica* 1987:204-211.

43. Panavas T, Rubinstein B: **Oxidative events during programmed cell death of daylily (*Hemerocallis hybrid*) petals**. *Plant Science* 1998, **133**:125 - 138.

44. Fukuchi-mizutani M, Ishiguro K, Nakayama T: **Molecular and functional characterization of a rose lipoxygenase cDNA related to flower senescence**. *Plant Cell* 2000, **160**:129 - 137.

45. Alba R, Fei Z, Payton P, et al. **ESTs, cDNA microarrays, and gene expression profiling: tools for dissecting plant physiology and development.** *The Plant journal : for cell and molecular biology* 2004, **39**:697-714.

46. Meyers BC, Galbraith DW, Nelson T, Agrawal V: **Methods for transcriptional profiling in plants. Be fruitful and replicate**. *Plant Physiology* 2004, **135**:637.

47. Adams MD, Kelley JM, Gocayne JD, et al. **Complementary DNA sequencing: expressed sequence tags and human genome project** . *Science* 1991, **252** :1651-1656.

48. Liang P, Pardee AB: **Differential Display**. *Differential Display Methods and Protocols* 1997, **85**:3-11.

49. Velculescu VE, Zhang L, Vogelstein B, Kinzler KW: **Serial analysis of gene expression.** *Science (New York, N.Y.)* 1995, **270**:484-7.

50. Bachem CW, Hoeven RS van der, Bruijn SM de, et al. **Visualization of differential gene expression using a novel method of RNA fingerprinting based on AFLP: analysis of gene expression during potato tuber development.** *The Plant journal : for cell and molecular biology* 1996, **9**:745-53.

51. others: **AFLP: a new technique for DNA fingerprinting**. *Nucleic acids research* 1995, **23**:4407.

52. Diatchenko L, Laut Y-fai C, Campbellt AP, et al. **differentially regulated or tissue-specific cDNA probes**. *Proc. Natl. Acad. Sci.* 1996, **93**:6025-6030.

53. others: **Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays**. *Nature biotechnology* 2000, **18**:630–634.

54. Sterky F, Bhalerao RR, Unneberg P, et al. **A Populus EST resource for plant functional genomics**. *PNAS* 2004, **101**:13951-13956.

55. Allona I, Quinn M, Shoop E, et al. **Analysis of xylem formation in pine by cDNA sequencing**. *Plant biology* 1998, **95**:9693-9698.

56. Vinod MS, Sankararamasubramanian HM, Priyanka R, Ganesan G, Parida A: **Gene expression analysis of volatile-rich male flowers of dioecious *Pandanus fascicularis* using expressed sequence tags.** *Journal of plant physiology* 2010, **167**:914-9.

57. Moccia MD, Oger-Desfeux C, Marais GA, Widmer A: **A White Campion (*Silene latifolia*) floral expressed sequence tag (EST) library: annotation, EST-SSR characterization, transferability, and utility for comparative mapping.** *BMC genomics* 2009, **10**:243.

58. Shimamura K, Ishimizu T, Nishimura K, et al. **Analysis of expressed sequence tags from Petunia flowers**. *Plant Science* 2007, **173**:495-500.

59. Shaogui G, Yi Z, Je-Gun J, et al. **Transcriptome sequencing and comparative analysis of cucumber flowers with different sex types**. *BMC Genomics* 2010, **11**.

60. Laitinen R A E, Immanen J, Auvinen P, et al.: **Analysis of the floral transcriptome uncovers new regulators of organ determination and gene families related to flower organ differentiation in *Gerbera hybrida* (Asteraceae)**. *Genome research* 2005, **15**:475.

61. Tsai W-C, Hsiao Y-Y, Lee S-H, et al. **Expression analysis of the ESTs derived from the flower buds of *Phalaenopsis equestris***. *Plant Science* 2006, **170**:426-432.

62. Beulé T, Camps C, Debiesse S, et al. **Transcriptome analysis reveals differentially expressed genes associated with the mantled homeotic flowering abnormality in oil palm (*Elaeis guineensis*)**. *Tree Genetics & Genomes* 2010, **7**:169-182.

63. Wu Q, Song J, Sun Y, et al. **Transcript profiles of *Panax quinquefolius* from flower, leaf and root bring new insights into genes related to ginsenosides biosynthesis and transcriptional regulation.** *Physiologia plantarum* 2010, **138**:134-49.

64. Conesa A, Götz S, García-Gómez JM, et al. **Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research.** *Bioinformatics (Oxford, England)* 2005, **21**:3674-6.

65. Altschul SF, Madden TL, Schäffer a a, et al. **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucleic acids research* 1997, **25**:3389-402.

66. Hunter S, Apweiler R, Attwood TK, et al. **InterPro: the integrative protein signature database.** *Nucleic acids research* 2009, **37**:D211-5.

67. Finn RD, Mistry J, Tate J, et al. **The Pfam protein families database.** *Nucleic acids research* 2010, **38**:D211-22.

68. Charbonnel-Campaa L, Lauga B, Combes D: **Isolation of a type 2 metallothionein-like gene preferentially expressed in the tapetum in *Zea mays*.** *Gene* 2000, **254**:199-208.

69. Hunter DA, Steele BC, Reid MS: **Identification of genes associated with perianth senescence in Daffodil (*Narcissus pseudonarcissus* L. Dutch Master)**. *Plant Science* 2002, **163**:13–21.

70. Nordberg M: **Metallothioneins: historical review and state of knowledge.** *Talanta* 1998, **46**:243-54.

71. Miller JD, Arteca RN, Pell EJ: **Senescence-associated gene expression during ozone-induced leaf senescence in *Arabidopsis***. *Plant Physiology* 1999, **120**:1015.

72. Etscheid M, Klumper S, Riesner D: **Accumulation of a Metallothionein-like mRNA in Norway Spruce Under Environmental Stress**. *Journal of Phytopathology* 1999, **147**:207–213.

73. Davies C, Robinson SP: **Differential screening indicates a dramatic change in mRNA profiles during grape berry ripening. Cloning and characterization of cDNAs encoding putative cell wall and stress response proteins.** *Plant physiology* 2000, **122**:803-12.

74. Reid S: **Up-regulation of two cDNA clones encoding metallothionein-like proteins in apple fruit during cool storage**. *Physiologia Plantarum* 1997, **100**:183-189.

75. Jones CS, Davies HV, McNicol RJ, TAylor MA: **Cloning of three genes up-regulated in ripening raspberry fruit (*Rubus idaeus* cv. Glen Clova)**. *Journal of plant physiology* , **153**:643-648.

76. Buchanan-Wollaston V: **Isolation of cDNA clones for genes that are expressed during leaf senescence in *Brassica napus*. Identification of a gene encoding a senescence-specific metallothionein-like protein**. *Plant Physiology* 1994, **105**:839-846.

77. Buchanan-Wollaston V, Ainsworth C: **Leaf senescence in *Brassica napus*: cloning of senescence related genes by subtractive hybridisation**. *Plant Molecular Biology* , **33**:821-834.

78. Breeze E, Wagstaff C, Harrison E, et al. **Gene expression patterns to define stages of post-harvest senescence in *Alstroemeria* petals**. *Plant Biotechnology Journal* 2004, **2**:155-168.

79. Channelière S, Rivière S, Scalliet G, et al. **Analysis of gene expression in rose petals using expressed sequence tags.** *FEBS letters* 2002, **515**:35-8.

80. Swanson R, Clark T, Preuss D: **Expression profiling of *Arabidopsis* stigma tissue identifies stigma-specific genes**. *Sexual Plant Reproduction* , **18**:163-171.

81. Tung C, Dwyer K, Nasrallah M: **Genome-wide identification of genes expressed in *Arabidopsis* pistils specifically along the path of pollen tube growth**. *Plant* 2005, **138**:977-989.

82. Goldman M, Pezzotti M, Seurinck J, Mariani C: **Developmental expression of tobacco pistil-specific genes encoding novel extensin-like proteins**. *The Plant Cell Online* 1992, **4**:1041-51.

83. Yamane H, Lee S-J, Kim B-D, Tao R, Rose JKC: **A coupled yeast signal sequence trap and transient plant expression strategy to identify genes encoding secreted proteins from peach pistils.** *Journal of experimental botany* 2005, **56**:2229-38.

84. Quiapim AC, Brito MS, Bernardes L a S, et al. **Analysis of the *Nicotiana tabacum* stigma/style transcriptome reveals gene expression differences between wet and dry stigma species.** *Plant physiology* 2009, **149**:1211-30.

85. Allen MD, Yamasaki K, Ohme-Takagi M, Tateno M, Suzuki M: **A novel mode of DNA recognition by a beta-sheet revealed by the solution structure of the GCC-box binding domain in complex with DNA.** *The EMBO journal* 1998, **17**:5484-96.

86. Schopfer P: **Hydroxyl radical-induced cell-wall loosening in vitro and in vivo: implications for the control of elongation growth**. *The Plant Journal* 2001, **28**:679-688.

87. Brownleader MD, Hopkins J, Mobasheri A, et al. **Role of extensin peroxidase in tomato (*Lycopersicon esculentum* Mill.) seedling growth**. *Planta* , **210**:668-676.

88. Mudge SR, Rae AL, Diatloff E, Smith FW: **Expression analysis suggests novel roles for members of the Pht1 family of phosphate transporters in *Arabidopsis*.** *The Plant journal : for cell and molecular biology* 2002, **31**:341-53.

89. Chapin LJ, Jones ML: **Ethylene regulates phosphorus remobilization and expression of a phosphate transporter (PhPT1) during petunia corolla senescence.** *Journal of experimental botany* 2009, **60**:2179-90.

90. Schneider S, Schneidereit A, Udvardi P, et al. ***Arabidopsis* INOSITOL TRANSPORTER2 mediates H+ symport of different inositol epimers and derivatives across the plasma membrane.** *Plant physiology* 2007, **145**:1395-407.

91. Roberts J: **Cell Separation Processes in Plants—Models, Mechanisms and Manipulation**. *Annals of Botany* 2000, **86**:223-235.

92. Sexton R, Roberts JA: **Cell Biology of Abscission**. *Annual Review of Plant Physiology* 1982, **33**:133-162.

93. Abeles F: **Role of RNA and protein synthesis in abscission**. *Plant Physiology* 1968, **43**:1577.

94. Roberts J: **Cell Separation Processes in Plants—Models, Mechanisms and Manipulation**. *Annals of Botany* 2000, **86**:223-235.

95. Bonghi C, Rascio N, Ramina A, Casadoro G: **Cellulase and polygalacturonase involvement in the abscission of leaf and fruit explants of peach**. *Plant Molecular Biology* 1992, **20**:839-848.

96. Lashbrook CC, Gonzalez-Bosch C, Bennett a B: **Two divergent endo-beta-1,4-glucanase genes exhibit overlapping expression in ripening fruit and abscising flowers.** *The Plant cell* 1994, **6**:1485-93.

97. Kalaitzis P, Koehler SM, Tucker ML: **Cloning of a tomato polygalacturonase expressed in abscission**. *Plant Molecular Biology* 1995, **28**:647-656.

98. TAYLOR JE, WEBB STJ, COUPE SA, TUCKER GA, ROBERTS JA: **Changes in Polygalacturonase Activity and Solubility of Polyuronides during Ethylene-stimulated Leaf Abscission in *Sambucus nigra*** . *Journal of Experimental Botany* 1993, **44** :93-98.

99. Woollard A a D, Moore I: **The functions of Rab GTPases in plant membrane traffic.** *Current opinion in plant biology* 2008, **11**:610-9.

100. Vernoud V, Horton AC, Yang Z, Nielsen E: **Analysis of the Small GTPase Gene Superfamily of *Arabidopsis*** . *Plant Physiology* 2003, **131** :1191-1208.

101. Cheung AY, Chen CY, Tao L, et al. **Regulation of pollen tube growth by Rac-like GTPases** . *Journal of Experimental Botany* 2003, **54** :73-81.

102. Hwang J-U, Yang Z: **Small GTPases and Spatiotemporal Regulation of Pollen Tube Growth**. In *The Pollen Tube*. edited by Malhó R Springer Berlin / Heidelberg; 2006, **3**:95-116.

103. Kotchoni SO, Jimenez-Lopez JC, Gao D, et al. **Modeling-Dependent Protein Characterization of the Rice Aldehyde Dehydrogenase (ALDH) Superfamily Reveals Distinct Functional and Structural Features**. *PLoS ONE* 2010, **5**:e11516.

104. Wei Y, Lin M, Oliver DJ, Schnable PS: **The roles of aldehyde dehydrogenases (ALDHs) in the PDH bypass of *Arabidopsis*.** *BMC biochemistry* 2009, **10**:7.

105. Zielinski RE: **Calmodulin and calmodulin-binding proteins in plants**. *Annual Review of Plant Biology* 1998, **49**:697–725.

106. Arazi T, Baum G, Snedden WA, Shelp BJ, Fromm H: **Molecular and Biochemical Analysis of Calmodulin Interactions with the Calmodulin-Binding Domain of Plant Glutamate Decarboxylase** . *Plant Physiology* 1995, **108** :551-561.

107. Bachs O, Agell N, Carafoli E: **Calmodulin and calmodulin-binding proteins in the nucleus**. *Cell Calcium* 1994, **16**:289-296.

108. Raz V, Fluhr R: **Ethylene Signal Is Transduced via Protein Phosphorylation Events in Plants** . *The Plant Cell Online* 1993, **5** :523-530.

109. Abeles F: **Role of RNA and protein synthesis in abscission**. *Plant Physiology* 1968, **43**:1577.

110. Takatsuji H: **Zinc-finger transcription factors in plants.** *Cellular and molecular life sciences : CMLS* 1998, **54**:582-96.

111. Weigel D: **From floral induction to floral shape**. *Current Opinion in Plant Biology* 1998, **1**:55–59.

112. Hu W, Wang Y, Bowers C, Ma H: **Isolation, sequence analysis, and expression studies of florally expressed cDNAs in *Arabidopsis***. *Plant Molecular Biology* 2003, **53**:545-563.

113. Zhao D, Yu Q, Chen C, Ma H: **Genetic control of reproductive meristems**. *Sheffield Academic Press* 2001:89– 142.

114. Santos LA: **Efeito da Superexpressão dos Fatores de Transcrição ZmDof1 e OsDof25 sobre a Eficiência de Uso de Nitrogênio em *Arabidopsis thaliana* L.** *Carbon* 2009.

115. Zhang J: **Overexpression analysis of plant transcription factors**. *Current Opinion in Plant Biology* 2003, **6**:430-440.

116. Liu L, White MJ, Macrae TH: **Functional domains , evolution and regulation**. 1999, **257**.

117. Cero JD: **Análise "in silico" e de expressão da família gênica ETHYLENE RESPONSIVE FACTORS (ERF) no gênro Malus**. 2010.

118. Perez-Rodriguez P, Riano-Pachon D, Correa LGG, et al. **PlnTFDB: updated content and new features of the plant transcription factor database**. *Nucleic Acids Research* 2009.

119. Guo AY, Chen X, Gao G, et al. **PlantTFDB: a comprehensive plant transcription factor database**. *Nucleic Acids Research* 2008, **36**:D966.

120. Dietz K-J, Vogel MO, Viehhauser A: **AP2/EREBP transcription factors are part of gene regulatory networks and integrate metabolic, hormonal and environmental signals in stress acclimation and retrograde signalling.** *Protoplasma* 2010:3-14.

121. Feng J-X, Liu D, Pan Y, et al. **An annotation update via cDNA sequence analysis and comprehensive profiling of developmental, hormonal or environmental responsiveness of the *Arabidopsis* AP2/EREBP transcription factor gene family.** *Plant molecular biology* 2005, **59**:853-68.

122. Huibers RP, Jong M de, Dekter RW, Van den Ackerveken G: **Disease-specific expression of host genes during downy mildew infection of *Arabidopsis*.** *Molecular plant-microbe interactions : MPMI* 2009, **22**:1104-15.

123. Balazadeh S, Riaño-Pachón DM, Mueller-Roeber B: **Transcription factors regulating leaf senescence in *Arabidopsis thaliana*.** *Plant biology (Stuttgart, Germany)* 2008, **10 Suppl 1**:63-75.

124. Yamasaki K, Kigawa T, Inoue M, Tateno M, Yamasaki T: **Solution Structure of the B3 DNA Binding Domain of the *Arabidopsis* Cold-Responsive Transcription Factor RAV1**. *Structure* 2004, **16**:3448-3459.

125. Kagaya Y, Ohmiya K, Hattori T: **RAV1, a novel DNA-binding protein, binds to bipartite recognition sequence through two distinct DNA-binding domains uniquely found in higher plants**. *Nucleic acids research* 1999, **27**:470.

126. Swaminathan K, Peterson K, Jack T: **The plant B3 superfamily.** *Trends in plant science* 2008, **13**:647-55.

127. GUO Y, CAI Z, GAN S: **Transcriptome of Arabidopsis leaf senescence**. *Plant, Cell & Environment* 2004, **27**:521-549.

128. Wellmer F, Riechmann JL, Alves-Ferreira M, Meyerowitz EM: **Genome-wide analysis of spatial gene expression in *Arabidopsis* flowers**. *Plant Cell* 2004, **16**:1314-1326.

129. Hennig L, Gruissem W, Grossniklaus U, Köhler C: **Transcriptional programs of early reproductive stages in *Arabidopsis***. *Plant Physiology* 2004, **135**:1765-1775.

130. Seki M, Narusaka M, Ishida J, et al. **Monitoring the expression profiles of 7000 *Arabidopsis* genes under drought, cold and high-salinity stresses using a full-length cDNA microarray**. *The Plant Journal* 2002, **31**:279-292.

131. Hoth S, Morgante M, Sanchez J-P, et al. **Genome-wide gene expression profiling in *Arabidopsis thaliana* reveals new targets of abscisic acid and largely impaired gene regulation in the abi1-1 mutant** . *Journal of Cell Science* 2002, **115** :4891-4900.

132. Schenk PM, Kazan K, Manners JM, et al. **Systemic gene expression in *Arabidopsis* during an incompatible interaction with Alternaria brassicicola**. *Plant Physiology* 2003, **132**:999-1010.

133. Oono Y, Seki M, Nanjo T, et al. **Monitoring expression profiles of *Arabidopsis* gene expression during rehydration process after dehydration using ca. 7000 full-length cDNA microarray**. *The Plant Journal* 2003, **34**:868-887.

134. Rabbani MA, Maruyama K, Abe H, et al. **Monitoring Expression Profiles of Rice Genes under Cold, Drought, and High-Salinity Stresses and Abscisic Acid Application Using cDNA Microarray and RNA Gel-Blot Analyses** . *Plant Physiology* 2003, **133** :1755-1767.

135. Deslandes L, Olivier J, Theulières F, et al. **Resistance to *Ralstonia solanacearum* in Arabidopsis thaliana is conferred by the recessive RRS1-R gene, a member of a novel family of resistance genes** . *Proceedings of the National Academy of Sciences* 2002, **99** :2404-2409.

136. Dellagi A, Heilbronn J, Avrova AO, et al. **A Potato Gene Encoding a WRKY-like Transcription Factor Is Induced in Interactions with *Erwinia carotovora* subsp. atroseptica and Phytophthora infestans and Is Coregulated with Class I Endochitinase Expression**. *Molecular Plant-Microbe Interactions* 2000, **13**:1092-1101.

137. Dong J, Chen C, Chen Z: **Expression profiles of the *Arabidopsis* WRKY gene superfamily during plant defense response**. *Plant Molecular Biology* , **51**:21-37.

138. Du L, Chen Z: **Identification of genes encoding receptor-like protein kinases as possible targets of pathogen- and salicylic acid-induced WRKY DNA-binding proteins in *Arabidopsis***. *The Plant Journal* 2000, **24**:837-847.

139. Chen C, Chen Z: **Potentiation of Developmentally Regulated Plant Defense Response by AtWRKY18, a Pathogen-Induced *Arabidopsis* Transcription Factor** . *Plant Physiology* 2002, **129** :706-716.

140. Chen W, Provart NJ, Glazebrook J, et al. **Expression Profile Matrix of *Arabidopsis* Transcription Factor Genes Suggests Their Putative Functions in Response to Environmental Stresses** . *The Plant Cell Online* 2002, **14** :559-574.

141. Wang Z, Yang P, Fan B, Chen Z: **An oligo selection procedure for identification of sequence-specific DNA-binding activities associated with the plant defence response**. *The Plant Journal* 1998, **16**:515-522.

142. Yang P, Chen C, Wang Z, Fan B, Chen Z: **A pathogen- and salicylic acid-induced WRKY DNA-binding activity recognizes the elicitor response element of the tobacco class I chitinase gene promoter**. *The Plant Journal* 1999, **18**:141-149.

143. Beyer K, Binder A, Boller T, Collinge M: **Identification of potato genes induced during colonization by *Phytophthora infestans***. *Molecular Plant Pathology* 2001, **2**:125-134.

144. Kalde M, Barth M, Somssich IE, Lippok B: **Members of the *Arabidopsis* WRKY Group III Transcription Factors Are Part of Different Plant Defense Signaling Pathways**. *Molecular Plant-Microbe Interactions* 2003, **16**:295-305.

145. Cheong YH, Chang H-S, Gupta R, et al. **Transcriptional Profiling Reveals Novel Interactions between Wounding, Pathogen, Abiotic Stress, and Hormonal Responses in *Arabidopsis*** . *Plant Physiology* 2002, **129** :661-677.

146. Rizhsky L, Liang H, Mittler R: **The Combined Effect of Drought Stress and Heat Shock on Gene Expression in Tobacco** . *Plant Physiology* 2002, **130** :1143-1151.

147. Pnueli L, Hallak-Herr E, Rozenberg M, et al. **Molecular and biochemical mechanisms associated with dormancy and drought tolerance in the desert legume *Retama raetam***. *The Plant Journal* 2002, **31**:319-330.

148. Huang T, Duman JG: **Cloning and characterization of a thermal hysteresis (antifreeze) protein with DNA-binding activity from winter bittersweet nightshade *Solanum dulcamara***. *Plant Molecular Biology* , **48**:339-350.

149. Johnson CS, Kolevski B, Smyth DR: **TRANSPARENT TESTA GLABRA2, a Trichome and Seed Coat Development Gene of *Arabidopsis*, Encodes a WRKY Transcription Factor** . *The Plant Cell Online* 2002, **14** :1359-1375.

150. Alexandrova KS, Conger BV: **Isolation of two somatic embryogenesis-related genes from orchardgrass (*Dactylis glomerata*)**. *Plant Science* 2002, **162**:301-307.

151. Robatzek S, Somssich IE: **A new member of the *Arabidopsis* WRKY transcription factor family, AtWRKY6, is associated with both senescence- and defence-related processes**. *The Plant Journal* 2001, **28**:123-133.

152. Robatzek S, Somssich IE: **Targets of AtWRKY6 regulation during plant senescence and pathogen defense.** *Genes & development* 2002, **16**:1139-49.

153. Hinderhofer K, Zentgraf U: **Identification of a transcription factor specifically expressed at the onset of leaf senescence**. *Planta* , **213**:469-473.

154. Rushton PJ, Macdonald H, Huttly AK, Lazarus CM, Hooley R: **Members of a new family of DNA-binding proteins bind to a conserved cis-element in the promoters of α-Amy2 genes**. *Plant Molecular Biology* , **29**:691-702.

155. Sun C, Palmqvist S, Olsson H, et al. **A Novel WRKY Transcription Factor, SUSIBA2, Participates in Sugar Signaling in Barley by Binding to the Sugar-Responsive Elements of the iso1 Promoter** . *The Plant Cell Online* 2003, **15** :2076-2092.

156. Willmott RL, Rushton PJ, Hooley R, Lazarus CM: **DNase1 footprints suggest the involvement of at least three types of transcription factors in the regulation of α-Amy2/A by gibberellin**. *Plant Molecular Biology* , **38**:817-825.

157. Kawaoka A, Kaothien P, Yoshida K, et al. **Functional analysis of tobacco LIM protein Ntlim1 involved in lignin biosynthesis**. *The Plant Journal* 2000, **22**:289-301.

158. Mundel C, Baltz R, Eliasson Å, et al. **A LIM-domain protein from sunflower is localized to the cytoplasm and/or nucleus in a wide variety of tissues and is associated with the phragmoplast in dividing cells**. *Plant Molecular Biology* , **42**:291-302.

159. Baltz R, Domon C, Pillay DT, Steinmetz A: **Characterization of a pollen-specific cDNA from sunflower encoding a zinc finger protein.** *The Plant Journal* 1992, **2**:713-721.

160. Eliasson Å, Gass N, Mundel C, et al. **Molecular and expression analysis of a LIM protein gene family from flowering plants**. *Molecular and General Genetics MGG* , **264**:257-267.

161. Feuerstein R, Wang X, Song D, Cooke NE, Liebhaber S a: **The LIM/double zinc-finger motif functions as a protein dimerization domain.** *Proceedings of the National Academy of Sciences of the United States of America* 1994, **91**:10655-9.

162. Johnson JD, Zhang W, Rudnick a, Rutter WJ, German MS: **Transcriptional synergy between LIM-homeodomain proteins and basic helix-loop-helix proteins: the LIM2 domain determines specificity.** *Molecular and cellular biology* 1997, **17**:3488-96.

163. Kong Y, Flick MJ, Kudla a J, Konieczny SF: **Muscle LIM protein promotes myogenesis by enhancing the activity of MyoD.** *Molecular and cellular biology* 1997, **17**:4750-60.

164. Kuroda S, Tokunaga C, Kiyohara Y, et al. **Protein-protein interaction of zinc finger LIM domains with protein kinase C.** *The Journal of biological chemistry* 1996, **271**:31029-32.

165. Jager M, Hassanin A, Manuel M, Le Guyader H, Deutsch J: **MADS-box genes in *Ginkgo biloba* and the evolution of the AGAMOUS family.** *Molecular biology and evolution* 2003, **20**:842-54.

166. Yanagisawa S: **Dof domain proteins: plant-specific transcription factors associated with diverse phenomena unique to plants**. *Plant Cell Physiology* 2004, **45**:386-391.

167. Zhang B, Chen W, Foley RC, Büttner M, Singh KB: **Interactions between distinct types of DNA binding proteins enhance binding to ocs element promoter sequences.** *The Plant cell* 1995, **7**:2241-52.

168. Kang HG, Singh KB: **Characterization of salicylic acid-responsive, arabidopsis Dof domain proteins: overexpression of OBP3 leads to growth defects.** *The Plant journal : for cell and molecular biology* 2000, **21**:329-39.

169. Yanagisawa S: **The Dof family of plant transcription factors**. *Trends in Plant Science* 2002, **7**:555-560.

170. Yanagisawa S, Sheen J: **Involvement of maize Dof zinc finger proteins in tissue-specific and light-regulated gene expression**. *The Plant Cell* 1998, **10**:75-89.

171. Yanagisawa S: **Dof1 and Dof2 transcription factors are associated with expression of multiple genes involved in carbon metabolism in maize**. *Plant J.* 2000, **21**:281-288.

172. Papi M, Sabatini S, Altamura MM, et al. **Inactivation of the phloem-specific Dof zinc finger gene DAG1 affects response to light and integrity of the testa of *Arabidopsis* seeds**. *Plant Physiol.* 2002, **128**:411-417.

173. Park DH, Lim PO, Kim JS, et al. **The *Arabidopsis* COG1 gene encodes a Dof domain transcription factor and negatively regulates phytochrome signaling.** *The Plant Journal* 2003, **34**:161-71.

174. De Paolis A, Sabatini S, De Pascalis L, Costantino P, Capone I: **A rolB regulatory factor belongs to a new class of single zinc finger plant proteins**. *The Plant Journal* 1996, **10**:215-223.

175. Kisu Y, Ono T, Shimofurutani N, Suzuki M, Esaka M: **Characterization and expression of a new class of zinc finger protein that binds to silencer region of ascorbate oxidase gene.** *Plant & cell physiology* 1998, **39**:1054-64.

176. Mena M, Cejudo FJ, Isabel-Lamoneda I, Carbonero P: **A role for the DOF transcription factor BPBF in the regulation of gibberellin-responsive genes in barley aleurone**. *Plant Physiol* 2002, **130**:111-119.

177. Washio K: **Identication of Dof proteins with implication in the gibberellin-regulated expression of a peptidase gene following the germination of rice grains**. *Biochemica et Biophysica Acta* 2001, **1520**:54-62.

178. Gualberti G, Papi M, Bellucci L, et al. **Mutations in the Dof zinc finger genes DAG2 and DAG1 influence with opposite effects the germination of *Arabidopsis* seeds**. *Plant Cell* 2002, **14**:1253-1263.

179. Papi M, Sabatini S, Bouchez D, et al. **Identification and disruption of an *Arabidopsis* zinc finger gene controlling seed germination**. *Genes & Development* 2000, **14**:28-33.

180. Mena M, Vicente-Carbajosa J, Schmidt RJ, Carbonero P: **An endosperm-specific DOF protein from barley, highly conserved in wheat, binds to and activates transcription from the prolamin-box of a native B-hordein promoter in barley endosperm**. *Plant J.* 1998, **16**:53-62.

181. Plesch G, Ehrhardt T, Mueller-roeber B: **Involvement of TAAAG elements suggests a role for Dof transcription factors in guard cell-speci®c gene expression**. *The Plant Journal* 2001, **28**:455-464.

182. Gan S, Amasino RM: **Making Sense of Senescence (Molecular Genetic Regulation and Manipulation of Leaf Senescence).** *Plant physiology* 1997, **113**:313-319.

183. Ling H: **Sequence analysis of GDSL lipase gene family in *Arabidopsis thaliana***. *Pak. J. Biol. Sci* 2008, **11**:763–767.

184. Kim WY, Hicks KA, Somers DE: **Independent roles for EARLY FLOWERING 3 and ZEITLUPE in the control of circadian timing, hypocotyl length, and flowering time**. *Plant physiology* 2005, **139**:1557.

185. Nakamura T, Song I-J, Fukuda T, et al. **Characterization of TrcMADS1 gene of *Trillium camtschatcense* (Trilliaceae) reveals functional evolution of the SOC1/TM3-like gene family**. *Journal of Plant Research* , **118**:229-234.

186. Ferrario S, Busscher J, Franken J, et al. **Ectopic Expression of the Petunia MADS Box Gene UNSHAVEN Accelerates Flowering and Confers Leaf-Like Characteristics to Floral Organs in a Dominant-Negative Manner** . *The Plant Cell Online* 2004, **16** :1490-1505.

187. Lee H: **The AGAMOUS-LIKE 20 MADS domain protein integrates floral inductive pathways in *Arabidopsis***. *Genes & Development* 2000, **14**:2366-2376.

188. Lee S, Kim J, Han J-J, Han M-J, An G: **Functional analyses of the flowering time gene OsMADS50, the putative SUPPRESSOR OF OVEREXPRESSION OF CO 1/AGAMOUS-LIKE 20 (SOC1/AGL20) ortholog in rice**. *The Plant Journal* 2004, **38**:754-764.

189. Lee J, Oh M, Park H, Lee I: **SOC1 translocated to the nucleus by interaction with AGL24 directly regulates LEAFY**. *The Plant Journal* 2008, **55**:832-843.

190. Liu C, Zhou J, Bracha-Drori K, et al. **Specification of *Arabidopsis* floral meristem identity by repression of flowering time genes** . *Development* 2007, **134** :1901-1910.

191. Liu C, Xi W, Shen L, Tan C, Yu H: **Regulation of Floral Patterning by Flowering Time Genes**. *Developmental Cell* 2009, **16**:711-722.

192. Melzer S, Lens F, Gennen J, et al. **Flowering-time genes modulate meristem determinacy and growth form in *Arabidopsis thaliana***. *Nat Genet* 2008, **40**:1489-1492.

193. Lee J, Lee I: **Regulation and function of SOC1, a flowering pathway integrator.** *Journal of experimental botany* 2010, **61**:2247-2254.

194. Livak KJ, Schmittgen TD: **Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method.** *Methods (San Diego, Calif.)* 2001, **25**:402-8.

195. ABRAF: **Anuário estatístico da ABRAF 2010 ano base 2009**. *ABRAF - Brasília* 2010, **54**:140p.

196. Century K, Reuber TL, Ratcliffe OJ: **Regulating the regulators: the future prospects for transcription-factor-based agricultural biotechnology products.** *Plant physiology* 2008, **147**:20-9.

197. Lijavetzky D, Carbonero P, Vicente-Carbajosa J: **Genome-wide comparative phylogenetic analysis of the rice and *Arabidopsis* Dof gene families.** *BMC evolutionary biology* 2003, **3**:17.

198. Yanagisawa S: **A novel DNA-binding zinc finger motif that may form a single zinc finger motif**. *Nucleic Acids Research* 1995, **23**:3403-3410.

200. Thornton J: **Gene family evolution and homology: genomics meet phylogenetics Ann. Rev**. *Genomics Human Genet* 2000.

201. Lomsadze A, Ter-Hovhannisyan V, Chernoff YO, Borodovsky M: **Gene identification in novel eukaryotic genomes by self-training algorithm.** *Nucleic acids research* 2005, **33**:6494-506.

202. Bateman A, Birney E, Cerruti L, et al. **The Pfam protein families database**. *Nucleic acids research* 2002, **30**:276.

203. Carver T, Berriman M, Tivey A, et al. **Artemis and ACT: viewing, annotating and comparing sequences stored in a relational database.** *Bioinformatics (Oxford, England)* 2008, **24**:2672-6.

204. Guo A-Y, Chen X, Gao G, et al. **PlantTFDB: a comprehensive plant transcription factor database.** *Nucleic acids research* 2008, **36**:D966-9.

205. Bailey TL, Williams N, Misleh C, Li WW: **MEME: discovering and analyzing DNA and protein sequence motifs.** *Nucleic acids research* 2006, **34**:W369-73.

206. Tamura K, Dudley J, Nei M, Kumar S: **MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0**. *Molecular biology and evolution* 2007, **24**:1596.

207. Czechowski T, Stitt M, Altmann T, Udvardi MK: **Genome-Wide Identification and Testing of Superior Reference Genes for Transcript Normalization**. *Plant Physiology* 2005, **139**:5-17.

208. Hruz T, Laule O, Szabo G, et al. **Genevestigator v3: a reference expression database for the meta-analysis of transcriptomes.** *Advances in bioinformatics* 2008, **2008**:420747.

209. Czechowski T, Bari RP, Stitt M, Scheible W-R, Udvardi MK: **Real-time RT-PCR profiling of over 1400 *Arabidopsis* transcription factors: unprecedented sensitivity reveals novel root- and shoot-specific genes.** *The Plant journal : for cell and molecular biology* 2004, **38**:366-79.

210. Holland MJ: **Transcript abundance in yeast varies over six orders of magnitude.** *The Journal of biological chemistry* 2002, **277**:14363-6.

211. Imaizumi T, Schultz TF, Harmon FG, Ho LA, Kay SA: **FKF1 F-Box Protein Mediates Cyclic Degradation of a Repressor of CONSTANS in *Arabidopsis*** . *Science* 2005, **309** :293-297.

212. Kim H-S, Kim SJ, Abbasi N, et al. **The DOF transcription factor Dof5.1 influences leaf axial patterning by promoting Revoluta transcription in Arabidopsis**. *The Plant Journal* 2010, **64**:524-535.

213. Gardiner J, Sherr I, Scarpella E: **Expression of Dof gnene idnetitifies eraly stages of vascular development in *Arabidopsis* leaves**. *The International Journal of Developmental Biology* 2010.

214. Wellmer F, Alves-Ferreira M, Dubois A, Riechmann JL, Meyerowitz EM: **Genome-Wide Analysis of Gene Expression during Early *Arabidopsis* Flower Development**. *PLoS Genetics* 2006, **2**:e117.

215. Wei P-C, Tan F, Gao X-Q, et al. **Overexpression of AtDOF4.7, an Arabidopsis DOF Family Transcription Factor, Induces Floral Organ Abscission Deficiency in *Arabidopsis***. *PLANT PHYSIOLOGY* 2010, **153**:1031-1045.

216. Derbyshire P, Drea S, Shaw PJ, Doonan JH, Dolan L: **Proximal-distal patterns of transcription factor gene expression during *Arabidopsis* root development.** *Journal of experimental botany* 2008, **59**:235-45.

217. Braam J, Davis RW: **Rain-, wind-, and touch-induced expression of calmodulin and calmodulin-related genes in *Arabidopsis***. *Cell* 1990, **60**:357-364.

218. Braam J, Sistrunk ML, Polisensky DH, et al. **Life in a changing world: TCH gene regulation of expression and responses to environmental signals**. *Physiologia Plantarum* 1996, **98**:909-916.

219. Braam J, Sistrunk ML, Polisensky DH, et al. **Plant responses to environmental stress: regulation and functions of the *Arabidopsis* TCH genes.** *Planta* 1997, **203 Suppl**:S35-41.

220. Yi H-S, Heil M, Adame-Alvarez RM, Ballhorn DJ, Ryu C-M: **Airborne induction and priming of plant defenses against a bacterial pathogen** . *Plant Physiology*

221. Swarbreck D, Wilks C, Lamesch P, et al. **The Arabidopsis Information Resource (TAIR): gene structure and function annotation.** *Nucleic acids research* 2008, **36**:D1009-14.

222. others: **The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray)**. *Science* 2006, **313**:1596.

223. Velasco R, Zharkikh A, Affourtit J, et al. **The genome of the domesticated apple (Malus [times] domestica Borkh.)**. *Nat Genet* 2010, **42**:833-839.

224. Jaillon O, Aury J-M, Noel B, et al. **The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla**. *Nature* 2007, **449**:463-467.

225. Mamiatis T, Fritsch EF, Sambrook J, Engel J: *Molecular Cloning A Laboratory Manual*. 1985, **5**:104-104.

226. Foucart C, Paux E, Ladouce N, et al. **Transcript profiling of a xylem vs phloem cDNA subtractive library identifies new genes expressed during xylogenesis in *Eucalyptus*.** *The New phytologist* 2006, **170**:739-52.

227. Grattapaglia D, Kirst M: ***Eucalyptus* applied genomics: from gene sequences to breeding tools.** *The New phytologist* 2008, **179**:911-29.

228. Steane DA, Nicolle D, McKinnon GE, Vaillancourt RE, Potts BM: **Higher-level relationships among the eucalypts are resolved by ITS-sequence data**. *Australian Systematic Botany* 2002, **15**:49-62.

229. Rengel D, San Clemente H, Servant F, et al. **A new genomic resource dedicated to wood formation in *Eucalyptus***. *BMC Plant Biology* 2009, **9**:36.

230. Kim S, Kim T: **Selection of optimal internal controls for gene expression profiling of liver disease.** *BioTechniques* 2003, **35**:456-8, 460.

231. Rajeevan MS, Vernon SD, Taysavang N, Unger ER: **Validation of array-based gene expression profiles by real-time (kinetic) RT-PCR.** *The Journal of molecular diagnostics : JMD* 2001, **3**:26-31.

232. Aerts JL, Gonzales MI, Topalian SL: **Selection of appropriate control genes to assess expression of tumor antigens using real-time RT-PCR.** *BioTechniques* 2004, **36**:84-6, 88, 90-1.

233. Barsalobres-Cavallari CF, Severino FE, Maluf MP, Maia IG: **Identification of suitable internal control genes for expression studies in Coffea arabica under different experimental conditions**. *BMC Molecular Biology* 2009, **10**:1.

234. Bas A, Forsberg G, Hammarström S, Hammarström M-L: **Utility of the housekeeping genes 18S rRNA, beta-actin and glyceraldehyde-3-phosphate-dehydrogenase for normalization in real-time quantitative reverse transcriptase-polymerase chain reaction analysis of gene expression in human T lymphocytes.** *Scandinavian journal of immunology* 2004, **59**:566-73.

235. Schmidt GW, Delaney SK: **Stable internal reference genes for normalization of real-time RT-PCR in tobacco (*Nicotiana tabacum*) during development and abiotic stress**. *Molecular Genetics and Genomics* 2010, **283**:233-41.

236. Yang Y, Hou S, Cui G, et al. **Characterization of reference genes for quantitative real-time PCR analysis in various tissues of *Salvia miltiorrhiza***. *Molecular Biology Reports* 2010, **37**:507-13.

237. Tong Z, Gao Z, Wang F, Zhou J, Zhang Z: **Selection of reliable reference genes for gene expression studies in peach using real-time PCR**. *BMC Molecular Biology* 2009, **10**:71.

238. Maroufi A, Van Bockstaele E, De Loose M: **Validation of reference genes for gene expression analysis in chicory (*Cichorium intybus*) using quantitative real-time PCR**. *BMC Molecular Biology* 2010, **11**:15.

239. Guénin S, Mauriat M, Pelloux J, et al. **Normalization of qRT-PCR data: the necessity of adopting a systematic, experimental conditions-specific, validation of references.** *Journal of experimental botany* 2009, **60**:487-93.

240. Libault M, Thibivilliers S, Bilgin DD, et al. **Identification of four soybean reference genes for gene expression normalization**. *The Plant Genome* 2008, **1**:44-54.

241. Silveira ED, Alves-Ferreira M, Guimarães LA, Silva FR da, Carneiro VTDC: **Selection of reference genes for quantitative real-time PCR expression studies in the apomictic and sexual grass *Brachiaria brizantha***. *BMC Plant Biology* 2009, **9**:84.

242. Hruz T, Wyss M, Docquier M, et al. **RefGenes: identification of reliable and condition specific reference genes for RT-qPCR data normalization**. *BMC Genomics* 2011, **12**:156.

243. Thellin O, Zorzi W, Lakaye B, et al. **Housekeeping genes as internal standards: use and limits**. *Journal of Biotechnology* 1999, **75**:291-295.

244. Jain M, Nijhawan A, Tyagi AK, Khurana JP: **Validation of housekeeping genes as internal control for studying gene expression in rice by quantitative real-time PCR**. *Biochemical and biophysical research communications* 2006, **345**:646-51.

245. Remans T, Smeets K, Opdenakker K, et al. **Normalisation of real-time RT-PCR gene expression measurements in *Arabidopsis thaliana* exposed to increased metal concentrations**. *Planta* 2008, **227**:1343-9.

246. Iskandar HM, Simpson RS, Casu RE, et al. **Comparison of reference genes for quantitative real-time polymerase chain reaction analysis of gene expression in sugarcane**. *Plant Molecular Biology Reporter* 2004, **22**:325-337.

247. Nicot N, Hausman JF, Hoffmann L, Evers D: **Housekeeping gene selection for real-time RT-PCR normalization in potato during biotic and abiotic stress**. *Journal of Experimental Botany* 2005, **56**:2907-2914.

248. Hong S-Y, Seo PJ, Yang M-S, Xiang F, Park C-M: **Exploring valid reference genes for gene expression studies in *Brachypodium distachyon* by real-time PCR.** *BMC plant biology* 2008, **8**:112.

249. Hu R, Fan C, Li H, Zhang Q, Fu Y-F: **Evaluation of putative reference genes for gene expression normalization in soybean by quantitative real-time RT-PCR**. *BMC Molecular Biology* 2009, **10**:93.

250. Jian B, Liu B, Bi Y, et al. **Validation of internal control for gene expression study in soybean by quantitative real-time PCR**. *BMC Molecular Biology* 2008, **9**:59.

251. Coker JS, Davies E: **Selection of candidate housekeeping controls in tomato plants using EST data.** *BioTechniques* 2003, **35**:740-2, 744, 746 passim.

252. Expósito-Rodríguez M, Borges AA, Borges-Pérez A, Pérez JA: **Selection of internal control genes for quantitative real-time RT-PCR studies during tomato development process**. *BMC Plant Biology* 2008, **8**:131.

253. Paolacci AR, Tanzarella OA, Porceddu E, Ciaffi M: **Identification and validation of reference genes for quantitative RT-PCR normalization in wheat.** *BMC molecular biology* 2009, **10**:11.

254. Artico S, Nardeli SM, Brilhante O, Grossi-de-Sa MF, Alves-Ferreira M: **Identification and evaluation of new reference genes in *Gossypium hirsutum* for accurate normalization of real-time quantitative RT-PCR data.** *BMC plant biology* 2010, **10**:49.

255. Wan H, Zhao Z, Qian C, et al. **Selection of appropriate reference genes for gene expression studies by quantitative real-time polymerase chain reaction in cucumber**. *Analytical Biochemistry* 2010, **399**:257-61.

256. Brunner AM, Yakovlev IA, Strauss SH: **Validating internal controls for quantitative plant gene expression studies**. *BMC Plant Biology* 2004, **4**:14.

257. Gutierrez L, Mauriat M, Guénin S, et al. **The lack of a systematic validation of reference genes: a serious pitfall undervalued in reverse transcription-polymerase chain reaction (RT-PCR) analysis in plants.** *Plant biotechnology journal* 2008, **6**:609-18.

258. Reid K, Olsson N, Schlosser J, Peng F, Lund S: **An optimized grapevine RNA isolation procedure and statistical determination of reference genes for real-time RT-PCR during berry development**. *BMC Plant Biology* 2006, **6**:27.

259. Lin YL, Lai ZX: **Reference gene selection for qPCR analysis during somatic embryogenesis in longan tree**. *Plant Science* 2010.

260. Almeida MR de, Ruedell CM, Ricachenevsky FK, et al. **Reference gene selection for quantitative reverse transcription-polymerase chain reaction normalization during in vitro adventitious rooting in *Eucalyptus globulus* Labill.** *BMC molecular biology* 2010, **11**:73.

261. Boava LP, Kuhn OJ, Pascholati SF, Di Piero RM, Luiz E: **Efeito de indutores bióticos e abióticos na atividade de quitinase e peroxidase e no controle da ferrugem causada por *Puccinia psidii* em eucalipto**. *SciELO Brasil*

262. Tusher VG, Tibshirani R, Chu G: **Significance analysis of microarrays applied to the ionizing radiation response.** *Proceedings of the National Academy of Sciences of the United States of America* 2001, **98**:5116-21.

263. Dambrauskas G, Aves SJ, Bryant JA, Francis D, Rogers HJ: **Genes encoding two essential DNA replication activation proteins, Cdc6 and Mcm3, exhibit very different patterns of expression in the tobacco BY-2 cell cycle.** *Journal of experimental botany* 2003, **54**:699-706.

264. Vandesompele J, De Preter K, Pattyn F, et al. **Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes**. *Genome Biology* 2002, **3**:1-12.

265. Andersen CL, Jensen JL, Ørntoft TF: **Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited**

for normalization, applied to bladder and colon cancer data sets. *Cancer research* 2004, **64**:5245-50.

266. Allen D, Winters E, Kenna PF, Humphries P, Farrar GJ: **Reference gene selection for real-time rtPCR in human epidermal keratinocytes.** *Journal of dermatological science* 2008, **49**:217-25.

267. Coll A, Nadal A, Collado R, et al. **Natural variation explains most transcriptomic changes among maize plants of MON810 and comparable non-GM varieties subjected to two N-fertilization farming practices.** *Plant molecular biology* 2010, **73**:349-62.

268. Radonić A, Thulke S, Bae H-G, et al. **Reference gene selection for quantitative real-time PCR analysis in virus infected cells: SARS corona virus, Yellow fever virus, Human Herpesvirus-6, Camelpox virus and Cytomegalovirus infections.** *Virology journal* 2005, **2**:7.

269. Taylor S, Wakem M, Dijkman G, Alsarraj M, Nguyen M: **A practical approach to RT-qPCR-Publishing data that conform to the MIQE guidelines.** *Methods (San Diego, Calif.)* 2010, **50**:S1-5.

270. Wu SJ, Shi M, Wu JY: **Cloning and characterization of the 1-deoxy-D-xylulose 5-phosphate reductoisomerase gene for diterpenoid tanshinone biosynthesis in *Salvia miltiorrhiza* (Chinese sage) hairy roots.** *Biotechnology and applied biochemistry* 2009, **52**:89-95.

271. Carretero-Paulet L, Cunillera N, Rodrı M, et al. **Expression and Molecular Analysis of the *Arabidopsis* DXR Gene Encoding 1-Deoxy- D -Xylulose 5-Phosphate Reductoisomerase , the First Committed Enzyme of the 2-C-Methyl- D -Erythritol 4-Phosphate Pathway 1**. *Plant Physiology* 2002, **129**:1581-1591.

272. Liao Z, Chen R, Min C, et al. **A New 1-Deoxy-D-xylulose 5-Phosphate Reductoisomerase Gene Encoding the Committed-Step Enzyme in the MEP Pathway from *Rauvolfia verticillata***. *Z. Naturforsch* 2007, **62c**:296-304.

273. Yan X, Zhang L, Wang J, et al. **Molecular characterization and expression of 1-deoxy-D-xylulose 5-phosphate reductoisomerase (DXR) gene from *Salvia miltiorrhiza***. *Acta Physiologiae Plantarum* , **31**:1015-1022.

274. Yue H, Eastman PS, Wang BB, et al. **An evaluation of the performance of cDNA microarrays for detecting changes in global mRNA expression.** *Nucleic acids research* 2001, **29**:E41-1.

275. Stankovic KM, Corfas G: **Real-time quantitative RT-PCR for low-abundance transcripts in the inner ear : analysis of neurotrophic factor expression**. *Hearing Research* 2003, **185**:97-108.

276. Stahlberg A, Hakansson J, Xian X, Semb H, Kubista M: **Properties of the reverse transcription reaction in mRNA quantification**. *Clinical chemistry* 2004, **50**:509.

277. Bustin S a: **Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): trends and problems.** *Journal of molecular endocrinology* 2002, **29**:23-39.

278. Gachon C, Mingam A, Charrier B: **Real-time PCR: what relevance to plant studies?** *Journal of experimental botany* 2004, **55**:1445-54.

279. Higuchi R, Fockler C, Dollinger G: **Kinetic PCR analysis: real-time monitoring of DNA amplification reactions**. *Nature Biotechnology* 1993, **11**:1026–1030.

280. Chuaqui RF, Bonner RF, Best CJM, et al. **Post-analysis follow-up and validation of microarray experiments.** *Nature genetics* 2002, **32 Suppl**:509-14.

281. Giulietti a, Overbergh L, Valckx D, et al. **An overview of real-time quantitative PCR: applications to quantify cytokine gene expression.** *Methods (San Diego, Calif.)* 2001, **25**:386-401.

282. Czechowski T, Bari RP, Stitt M, Scheible W-R, Udvardi MK: **Real-time RT-PCR profiling of over 1400 *Arabidopsis* transcription factors: unprecedented sensitivity reveals novel root- and shoot-specific genes.** *The Plant journal : for cell and molecular biology* 2004, **38**:366-79.

283. Gutierrez L, Mauriat M, Pelloux J, Bellini C, Van Wuytswinkel O: **Towards a systematic validation of references in real-time rt-PCR.** *The Plant cell* 2008, **20**:1734-5.

284. Volkov R a, Panchuk II, Schöffl F: **Heat-stress-dependency and developmental modulation of gene expression: the potential of house-keeping genes as internal standards in mRNA expression profiling using real-time RT-PCR.** *Journal of experimental botany* 2003, **54**:2343-9.

285. Martin RC, Hollenbeck VG, Dombrowski JE: **Evaluation of Reference Genes for Quantitative RT-PCR in *Lolium perenne***. *Crop Science* 2008, **48**:1881-1887.

286. Hays D: **Development of an expressed sequence tag (EST) library for Medicago sativa**. *Plant Science* 2001, **161**:517-526.

287. Sterky F, Regan S, Karlsson J, et al. **Gene discovery in the wood-forming tissues of poplar: analysis of 5, 692 expressed sequence tags.** *Proceedings of the National Academy of Sciences of the United States of America* 1998, **95**:13330-5.

288. González-Verdejo CI, Die JV, Nadal S, et al. **Selection of housekeeping genes for normalization by real-time RT-PCR: analysis of Or-MYB1 gene expression in *Orobanche ramosa* development.** *Analytical biochemistry* 2008, **379**:176-81.

289. Hong SM, Bahn SC, Lyu A, Jung HS, Ahn JH: **Identification and Testing of Superior Reference Genes for a Starting Pool of Transcript Normalization in *Arabidopsis*.** *Plant & cell physiology* 2010, **51**:1694-1706.

290. Zhou L, Lim Q-E, Wan G, Too H-P: **Normalization with genes encoding ribosomal proteins but not GAPDH provides an accurate quantification of gene expressions in neuronal differentiation of PC12 cells.** *BMC genomics* 2010, **11**:75.

291. Kim B-R, Nam H-Y, Kim S-U, Kim S-I, Chang Y-J: **Normalization of reverse transcription quantitative-PCR with housekeeping genes in rice.** *Biotechnology letters* 2003, **25**:1869-72.

292. Løvdal T, Lillo C: **Reference gene selection for quantitative real-time PCR normalization in tomato subjected to nitrogen, cold, and light stress**. *Analytical Biochemistry* 2009, **387**:238-242.

293. McIntosh KB, Bonham-Smith PC: **The two ribosomal protein L23A genes are differentially transcribed in *Arabidopsis thaliana***. *Genome* 2005, **48**:443–454.

294. Wool IG: **Extraribosomal functions of ribosomal proteins**. *Trends in biochemical sciences* 1996, **21**:164–165.

295. Taylor MA, Arif SA, Pearce SR, et al. **Differential expression and sequence analysis of ribosomal protein genes induced in stolon tips of potato (*Solanum tuberosum* L.) during the early stages of tuberization**. *Plant physiology* 1992, **100**:1171.

296. Williams ME, Sussex IM: **Developmental regulation of ribosomal protein L16 genes in *Arabidopsis thaliana***. *The Plant Journal* 1995, **8**:65–76.

297. Revenkova E, Masson J, Koncz C, et al. **Involvement of *Arabidopsis thaliana* ribosomal protein S27 in mRNA degradation triggered by genotoxic stress.** *The EMBO journal* 1999, **18**:490-9.

298. Kim K-Y, Park S-W, Chung Y-S, et al. **Molecular cloning of low-temperature-inducible ribosomal proteins from soybean.** *Journal of experimental botany* 2004, **55**:1153-5.

299. Saez-Vasquez J, Gallois P, Delseny M: **Accumulation and nuclear targeting of BnC24, a *Brassica napus* ribosomal protein corresponding to a mRNA accumulating in response to cold treatment**. *Plant science* 2000, **156**:35–46.

300. De Angelis M, Di Cagno R, Huet C, et al. **Heat shock response in *Lactobacillus plantarum***. *Applied and environmental microbiology* 2004, **70**:1336.

301. Zhang Y, Berger S a: **Increased calcium influx and ribosomal content correlate with resistance to endoplasmic reticulum stress-induced cell death in mutant leukemia cell lines.** *The Journal of biological chemistry* 2004, **279**:6507-16.

302. Marty I, Brugidou C, Chartier Y, Meyer Y: **Growth-related gene expression in *Nicotiana tabacum* mesophyll protoplasts**. *The Plant Journal* 1993, **4**:265–278.

303. Moran DL: **Characterization of the structure and expression of a highly conserved ribosomal protein gene, L9, from pea.** *Gene* 2000, **253**:19-29.

304. Bastolla FM: **Seleção e avaliação de genes de referência para estudos de expressão gênica em *Eucalyptus***. 2007:http://www.lume.ufrgs.br/handle/10183/14358.

305. Menges M, Jager SM de, Gruissem W, Murray JAH: **Global analysis of the core cell cycle regulators of *Arabidopsis* identifies novel genes, reveals multiple and highly specific profiles of expression and provides a coherent model for plant cell cycle control.** *The Plant journal : for cell and molecular biology* 2005, **41**:546-66.

306. Joubès J, Chevalier C, Dudits D, et al. **CDK-related protein kinases in plants.** *Plant molecular biology* 2000, **43**:607-20.

307. Stone JM, Walker JC: **Plant protein kinase families and signal transduction.** *Plant physiology* 1995, **108**:451-7.

308. Umeda M, Shimotohno A, Yamaguchi M: **Control of cell division and transcription by cyclin-dependent kinase-activating kinases in plants.** *Plant & cell physiology* 2005, **46**:1437-42.

309. Reines D: **Elongation factor-dependent transcript shortening by template-engaged RNA polymerase II.** *The Journal of biological chemistry* 1992, **267**:3795-800.

310. Duchêne A-M, Giritch A, Hoffmann B, et al. **Dual targeting is the rule for organellar aminoacyl-tRNA synthetases in *Arabidopsis thaliana*.** *Proceedings of the National Academy of Sciences of the United States of America* 2005, **102**:16484-9.

311. Hoenemann C, Hohe A: **Selection of reference genes for normalization of quantitative real-time PCR in cell cultures of *Cyclamen persicum***. *Electronic Journal of Biotechnology* 2011, **14**.

312. Kamada T, Nito K, Hayashi H, et al. **Functional differentiation of peroxisomes revealed by expression profiles of peroxisomal genes in *Arabidopsis thaliana*.** *Plant & cell physiology* 2003, **44**:1275-89.