

Universidade Federal do Rio Grande do Sul

Instituto de Ciências Básicas da Saúde

Curso de Graduação em Biomedicina

Trabalho de Conclusão de Curso

Genômica Comparativa em *Eucalyptus grandis* Relativa a Genes Codificadores  
de Fatores de Transcrição da Família Dof

Gabriel Silveira D'Almeida

Orientação

Prof. Dr. Giancarlo Pasquali

M.Sc. Michèle Claire Bretton

2010

## Agradecimentos

Aos meus pais, pelo apoio e amor incondicionais.

Aos meus queridos colegas do curso de Biomedicina pelos anos que passamos juntos. Muito obrigado Cícero, Conrado, Mariane, Andressa, Marina, Susana, Káren, Carla e Daniel (os componentes da "turma dentro da turma").

Ao meu orientador Gian, pelas oportunidades e ensinamentos.

Ao pessoal do Laboratório de Biologia Molecular Vegetal, especialmente à Michèle, cuja (enorme) paciência e orientações foram cruciais para a realização deste trabalho (muito obrigado!).

Aos habitantes da Alameda Picles (mesmo os eventuais), pelas jantãs e discussões filosóficas que serviram pra animar as semanas, especialmente aos Srs. Von e Peters.

À Caroline, sempre presente no Gmail pra conversar e descontraír nos momentos de desespero (e foram muitos).

Ao Boschi, companheiro pra qualquer empreitada desde a pré-história (ou quarta série, tanto faz).

Ao Marcelo, companheiro certo pra jogar os últimos lançamentos do PS3, e futuro aluno da UFRGS.

Aos meus primos Alice e Lourenço, pelos finais de semana de diversão em Charqueadas.

Ao pessoal do laboratório de Biofísica, que me acolheu por um ano, especialmente meus orientadores Iuri e Tiago.

Ao pessoal do laboratório de Farmacologia, que me aturou por seis meses. Obrigado Ana, Vivi e Elaine.

## Índice Geral

Resumo .....	5
Introdução Compreensiva.....	7
A Família Dof de Fatores de Transcrição.....	8
Estrutura das Proteínas Dof.....	8
Funções dos Fatores de Transcrição Dof .....	9
<i>Eucalyptus grandis</i> como Alvo de Biotecnologia .....	10
Referências.....	13
COMPARATIVE GENOMICS OF THE <i>DOF</i> TYPE GENE FAMILY IN <i>EUCALYPTUS GRANDIS</i> <sup>1</sup> .....	15
FOOTNOTES .....	16
ABSTRACT.....	17
INTRODUCTION.....	18
RESULTS .....	20
Obtaining <i>Dof</i> Sequences from Databases .....	20
Alignment and Phylogenetic Analysis .....	20
DISCUSSION.....	23
MATERIALS AND METHODS.....	25
Database Search.....	25
Local BLAST Search and Sequence Confirmation.....	25
Alignment and Phylogenetic Analysis .....	25
LITERATURE CITED .....	27
FIGURE LEGENDS.....	29
Figure 1 .....	29
Figure 2 .....	29
Figure 3 .....	29
Figure 4 .....	29
Tables.....	30
Table I.....	30
Table II.....	32
Original Figure Files.....	34
Figure 1 .....	34
Figure 2 .....	35
Figure 3 .....	36
Figure 4 .....	37

Supplemental Material .....	39
Supplemental Text I .....	39
Conclusões e Perspectivas .....	42
Obtenção das Sequências Dof de <i>Eucalyptus grandis</i> .....	42
Realização do Trabalho .....	42
Análise da Expressão das Proteínas Dof .....	43

## Resumo

O objetivo que norteou o presente trabalho foi identificar as sequências gênicas correspondentes a fatores de transcrição da família Dof em *Eucalyptus grandis*, e realizar análises das mesmas, incluindo a análise filogenética. A família Dof ("*DNA binding with One Finger*") compreende fatores de transcrição exclusivos de plantas, caracterizados pela presença de um motivo de ligação ao DNA semelhante a "dedos-de-zinco". As proteínas Dof exercem efeitos ativadores ou repressores da transcrição, atuando sobre promotores de genes associados aos mais diferentes metabolismos vegetais. Inicialmente, foram obtidas as sequências de genes Dof dos vegetais *Arabidopsis thaliana*, *Arabidopsis lyrata*, *Carica papaya*, *Populus trichocarpa*, *Vitis vinifera*, *Sorghum bicolor*, *Chlamydomonas reinhardtii*, *Oryza sativa indica* e *Zea mays*, dos bancos de dados online "*The Arabidopsis Information resource*" (TAIR) e "*Plant Transcription Factor Database*". A sequência peptídica correspondente ao domínio Dof de cada um destes genes foi obtida por meio de alinhamentos utilizando a ferramenta online Pfam, do "*Sanger Institute*". As sequências peptídicas foram então reversamente traduzidas para nucleotídeos por meio da ferramenta online "*Expasy Reverse Translate*", e utilizadas em buscas de BLAST local no genoma de *E. grandis* gerado pelo projeto GENOLYPTUS. Arcabouços das sequências nucleotídicas Dof de altas identidades foram utilizados em buscas em diferentes bancos de dados online, incluindo aqueles disponíveis no GenBank, no *Populus EST Database*, no banco de dados do Projeto "Genolyptus: Rede Brasileira de Pesquisa do Genoma de *Eucalyptus*", e no banco de sequências genômicas de *E. grandis* "Eucagen", de acesso restrito. As 20 sequências de *E. grandis* resultantes foram confirmadas por meio de alinhamentos com as demais sequências Dof conhecidas.

As 20 sequências Dof obtidas foram traduzidas para aminoácidos e utilizadas em análises filogenéticas, incluindo alinhamentos com o programa MUSCLE. O alinhamento inicial demonstrou a existência de duas sequências idênticas, e uma sequência que não possui o domínio Dof completo. Sendo assim, as análises prosseguiram com 18 sequências Dof. Um novo alinhamento demonstrou que as sequências coincidiam perfeitamente com a descrição dos fatores de transcrição Dof, alinhando-se corretamente apenas no domínio Dof (altamente conservado), e variando consideravelmente fora do domínio Dof.

Foram construídas duas árvores filogenéticas pelo método de máxima verossimilhança, incluindo uma árvore contendo apenas as sequências mais recentes disponibilizadas no Eucagen, e outra incluindo as sequências Dof de *E. grandis* juntamente com as de *A. thaliana* e *P. trichocarpa*. Pelas árvores filogenéticas foi demonstrada a proximidade estrutural e possivelmente funcional dos fatores de transcrição Dof destas três espécies, classificando os genes em 10 grandes grupos de ortólogos, e definindo pares de parálogos que incluíram os genes recém-descobertos.

Cinquenta diferentes motivos conservados, além do domínio Dof, foram identificados com o emprego do programa MEME, sendo 31 destes compartilhados pelas três espécies vegetais analisadas. A distribuição de motivos conservados fora do domínio Dof reforçou a localização dos genes nas árvores filogenéticas, demonstrando que os supostos parálogos possuíam motivos em comum.

Portanto, foram obtidas 18 sequências completas de proteínas Dof derivadas do transcriptoma e confirmadas no genoma de *E. grandis*. Estas sequências possuem domínios de alta identidade com a descrição do domínio Dof, indicando suas funcionalidades, e possuem motivos conservados em comum com genes *Dof* de *A. thaliana* e *P. trichocarpa*. Ensaios visando a clonagem, a expressão heteróloga e ensaios de atividades, bem como estudos genéticos e transgênicos serão futuramente conduzidos com vistas a comprovar a função destes genes em *E. grandis*.

## Introdução Compreensiva

O controle da transcrição é considerado o fator mais importante da regulação gênica. Em procariotos, nos quais os genes estão organizados em operons e as células devem responder rapidamente a alterações no meio externo, o controle se dá por meio de proteínas repressoras, que precisam ser removidas para que a transcrição ocorra, ou proteínas ativadoras, que aumentam a taxa de formação de complexos de início da transcrição. Uma única enzima RNA polimerase está presente em células procarióticas, a qual é capaz de reconhecer e ligar motivos conservados do DNA por meio de sua subunidade variável sigma ( $\sigma$ ), e sintetizar o RNA sem o auxílio de outras enzimas. A modulação da transcrição de células procarióticas envolve eventualmente outros mecanismos de regulação, incluindo, entre outros, a própria tradução de proteínas (atenuação) (Zaha, A., Ferreira, H., Passaglia L. 2003).

Nos eucariotos, a maior parte dos genes está organizada em unidades de transcrição únicas, compostas por um gene e suas regiões controladoras a montante (promotor) e a jusante (terminador), classicamente. Nas células vegetais, estão presentes cinco tipos distintos de RNA polimerases, incluindo as RNA polimerases I, II e III (nucleares), a RNA polimerase mitocondrial e, também, as plastídicas como as dos cloroplastos. Os complexos proteicos basais (holoenzimas) das enzimas nucleares são incapazes de, por si só, reconhecer motivos na sequência do DNA e, conseqüentemente, de se ligar ao mesmo para promover a transcrição. Para que ocorra a transcrição, são necessários fatores acessórios de transcrição, isto é, moléculas proteicas que apresentam domínios de ligação ao DNA, e que são capazes de interagir com elementos regulatórios de ação *cis*. Os fatores de transcrição podem ser gerais, ligando-se a elementos *cis* presentes em quase todas as sequências promotoras como *TATA Box*, *CAAT Box* ou *CG Box*, ou específicos, capazes de reconhecer especificamente uma sequência *cis* dentro de cada grupo de promotores. Os fatores de transcrição específicos designam quais genes deverão ser ativados ou reprimidos em resposta a sinais endógenos ou exógenos. Eles determinam quando, onde e em resposta a que sinais os genes deverão ser transcritos ou reprimidos. A modulação da atividade dos próprios fatores de transcrição em resposta aos sinais é mediada por diferentes cascatas de ativação e transmissão. No caso da ativação da transcrição, os fatores específicos recrutam os fatores gerais e uma das DNA polimerases para que ocorra a transcrição. Portanto, as funções dos fatores de

transcrição incluem o reconhecimento e a ligação a motivos de regulação *cis*, o recrutamento (ou o bloqueio) de outros fatores de transcrição, o recrutamento (ativadores) ou o bloqueio (repressores) da RNA polimerase, o direcionamento da atividade da RNA polimerase e a iniciação ou o bloqueio da transcrição (Zaha, A., Ferreira, H., Passaglia L. 2003).

### **A Família Dof de Fatores de Transcrição**

A família Dof ("*DNA binding with One Finger*") compreende fatores de transcrição exclusivos de plantas, caracterizados pela presença de um motivo de ligação ao DNA semelhante aos "dedos-de-zinco". Estas proteínas podem agir tanto como ativadores quanto repressores, e parecem ter evoluído para controlar a expressão de genes de processos característicos de plantas (Yanagisawa, 2002). O primeiro gene *Dof* (MNB1/ZmDOF1) foi isolado a partir do milho, e está envolvido com outro gene *Dof* desta espécie (ZmDOF2) na regulação da expressão pela luz do gene que codifica a C4-fosfenolpiruvato-carboxilase. Ambos os genes são expressos constitutivamente em folhas e caules de milho e são, respectivamente, ativadores e repressores transcricionais do promotor deste gene em ensaios de expressão transiente em protoplastos de folhas de milho (Yanagisawa & Sheen, 1998). Nas últimas décadas, foram identificados fatores de transcrição Dof em diversas espécies de monocotiledôneas e dicotiledôneas, incluindo trigo, cevada, arroz, tabaco, *Arabidopsis*, abóbora, batata, ervilha e mamão (Yanagisawa, 2004). Os genes *Dof* parecem ter se originado de um ancestral comum, que manteve uma única cópia na alga unicelular *Chlamydomonas reinhardtii* e, posteriormente, expandiu-se para outros grupos taxonômicos, aumentando de número por meio de eventos de duplicação (Moreno-Risueno, Martínez, Vicente-Carbajosa, & Carbonero, 2007).

### **Estrutura das Proteínas Dof**

Na classificação da família Dof, a estrutura das proteínas está dividida em domínio Dof (região conservada N-terminal) e região "não-Dof" (região C-terminal altamente variável). O domínio Dof apresenta-se como um domínio de ligação ao DNA com 52 resíduos de aminoácidos, com uma estrutura única contendo quatro



resíduos de cisteína que ligam um átomo de zinco. O domínio Dof é designado pela sequência conservada CX<sub>2</sub>CX<sub>21</sub>CX<sub>2</sub>C, onde X corresponde a qualquer aminoácido. Este domínio permite aos fatores Dof reconhecerem motivos *cis* contendo a sequência 5'-AAAG-3' nos promotores de genes (Umemura, Ishiduka, Yamamoto, & Esaka, 2004; Yanagisawa & Schmidt, 1999). Já a região fora do domínio Dof não aparenta ter função de ligação ao DNA. Devido à sua alta variabilidade, a região não-Dof está provavelmente relacionada às diferentes funções exercidas por essas proteínas (Yanagisawa & Schmidt, 1999). Cada gene codifica apenas um domínio Dof e, conseqüentemente, cada proteína codificada apresenta apenas um motivo de ligação ao DNA (Lijavetzky, Carbonero, & Vicente-Carbajosa, 2003). Experimentos utilizando agentes quelantes de metal mostraram-se capazes de inibir a ligação dos fatores de transcrição Dof ao DNA. A adição posterior de zinco ao meio foi capaz de restaurar esta ligação. Além disto, a substituição dos resíduos conservados de cisteína por serina e alanina inibiu completamente a capacidade de ligação dessas proteínas ao DNA, *in vivo* e *in vitro* (Yanagisawa, 2004).

O domínio Dof também é caracterizado por ser um domínio bifuncional, capaz de mediar interações proteína-proteína. Interações deste tipo realizadas pelo domínio Dof já foram demonstradas em *Arabidopsis* e milho. Em ambos os casos, os fatores de transcrição Dof demonstraram ser capazes de se ligar a proteínas de resposta ao estresse contendo um domínio de zíper-de-leucinas. Também foi demonstrada a interação entre o domínio Dof e o domínio *HMG-Box*. Este domínio representa os sítios de ligação ao DNA de proteínas nucleares não-histonas da família do grupo de alta mobilidade (do inglês, *High Mobility Group*, ou HMG). Foi demonstrado que a ligação do fator de transcrição Dof1 de milho ao DNA foi facilitada pela proteína HMG. Considera-se que mesmo que apenas algumas proteínas Dof tenham apresentado interação proteína-proteína *in vitro*, a maioria desses fatores de transcrição deve ser capaz de realizar estas interações *in vivo*, tendo-se em vista que o domínio é altamente conservado em todos os indivíduos da família (Yanagisawa, 2004).

### **Funções dos Fatores de Transcrição Dof**

As funções dos produtos de genes *Dof* são tão variadas quanto suas regiões não-*Dof*. Já foram descritos fatores de transcrição *Dof* relacionados com o desenvolvimento do endosperma (Vicente-Carbajosa, Moose, Parsons, & Schmidt, 1997), metabolismo de carboidratos (Yanagisawa & Sheen, 1998), mecanismos de defesa específicos de vegetais (Baumann, De Paolis, Costantino, & Gualberti, 1999), germinação de sementes (Papi et al., 2000), florescimento (Fornara et al., 2009; Li et al., 2009), desenvolvimento vascular (Konishi & Yanagisawa, 2007) e resposta a fitormônios (Kisu, Ono, Shimofurutani, Suzuki, & Esaka, 1998). Além disso, proteínas *Dof* já foram incluídas em redes de interações de fatores de transcrição em resposta a estresses (Cooper et al., 2003), e ao silenciamento de certas enzimas, como a ascorbato-oxidase em abóbora (Kisu, Ono, Shimofurutani, Suzuki, & Esaka, 1998).

Os fatores de transcrição *Dof* podem ter sua ação modificada por alterações estruturais e a expressão dos genes *Dof* pode ser alterada pela própria via de metabolismo de um nutriente ou composto utilizado pela planta. O fator de transcrição TaDof1 de trigo está relacionado com o metabolismo de nitrogênio, sofrendo, inclusive, regulação positiva sob baixas concentrações deste composto (Kumar, Taware, Gaur, Guru, & Kumar, 2009). Alterações nas estruturas das proteínas *Dof* DAG1 e DAG2 de *Arabidopsis* determinaram modificações no processo de germinação da planta (Gualberti et al., 2002).

Portanto, os fatores de transcrição *Dof* estão envolvidos na regulação de genes característicos de plantas, cujos produtos podem possuir inúmeras funções. Visto que genes ortólogos normalmente apresentam funções semelhantes, estudos filogenéticos sugeriram que pistas sobre a função de fatores de transcrição *Dof* podem ser obtidas quando proteínas recém descobertas desta família são comparadas a outras já identificadas e devidamente caracterizadas (Lijavetzky, Carbonero, & Vicente-Carbajosa, 2003).

### ***Eucalyptus grandis* como Alvo de Biotecnologia**

O eucalipto é a arbórea mais cultivada do mundo para exploração comercial de madeira. No Brasil, o cultivo de eucalipto é voltado principalmente para a produção de celulose e papel que, além de abastecer o mercado interno, gera matéria-prima e produtos beneficiados para exportação (Medeiros, 2009). De acordo com

dados publicados pela Associação Brasileira de Celulose e Papel (BRACELPA), a exportação de celulose brasileira tem como principais mercados a Europa (43%), a China (26%), a América do Norte (18%), a Ásia/Oceania (12%) e a América Latina (1%). As vendas de celulose corresponderam a US\$ 1,5 bilhões apenas no primeiro quadrimestre de 2010. Já as exportações de papel têm como principais mercados a América Latina (52%), a Europa (21%), a América do Norte (10%), a África (6%), a Ásia/Oceania (7%) e a China (4%), e chegaram a US\$ 657 milhões neste mesmo período. Essa produção ocorreu em 220 empresas, de 450 municípios, em 17 estados, gerando 114 mil empregos diretos e 500 mil indiretos (BRACELPA, 2010 – dados acessáveis em <http://www.bracelpa.org.br/bra/estatisticas/pdf/booklet/booklet.pdf>).

A alta variabilidade característica do gênero *Eucalyptus*, com mais de 700 espécies descritas, despertou interesse no melhoramento genético destas plantas, visando modificações desejáveis para a indústria. No Brasil, *E. grandis* é a espécie mais importante para a obtenção de papel e pasta de celulose, principalmente na forma de híbridos *E. grandis* x *E. urophylla* (Colodette *et al.* 2004). Devido ao seu valor econômico para o Brasil, seu ótimo desenvolvimento em regiões de clima tropical e subtropical, seu rápido crescimento volumétrico, o bom conteúdo de celulose e a resistência a pragas, esta espécie tornou-se o candidato óbvio a estudos genéticos e genômicos. Em 2002, no âmbito do Fundo Setorial Verde-Amarelo do Ministério da Ciência e Tecnologia e com o apoio de 14 empresas de celulose e papel, foi lançado o projeto “GENOLYPTUS: Rede Brasileira de Pesquisa do Genoma de *Eucalyptus*”, pelo qual realizou-se o sequenciamento do transcriptoma de *E. grandis* e de outras espécies de *Eucalyptus*, além de vários outros estudos de genética, fitopatologia e fisiologia deste vegetal (Medeiros, 2009, & Jahns, 2008). Em virtude do impacto da cultura de *E. grandis* mundialmente, um projeto internacional de sequenciamento genômico foi sediado no *Joint Genome Institute* (JGI) do Departamento de Energia dos Estados Unidos (USDoE). A árvore selecionada para cedência do genoma foi desenvolvida pela empresa Suzano Papel e Celulose S.A. (Suzano, SP), e as amostras de ácidos nucléicos enviadas ao JGI foram preparadas na Universidade Federal de Viçosa e pelo nosso grupo no CBiot/UFRGS. O Projeto EucaGen, como foi denominado, conta com a coordenação dos Drs. Jerry Tuskan (JGI, USDoE, EUA), Zander Myburg (Universidade de Pretória, África do Sul) e Dario Grattapaglia (Embrapa-Cenargen, Brasília, DF). O banco de dados

genômicos do Projeto EucaGen é de acesso restrito, incluindo nosso grupo de pesquisa, e exibe uma cobertura de 8X atualmente.

A anotação da família Dof de fatores de transcrição a partir dos dados do genoma de *E. grandis* contribuirá para o enriquecimento de bancos de dados de fatores de transcrição utilizados por futuros estudos de melhoramento genético de diversas plantas.

## Referências

Baumann K, De Paolis A, Costantino P, Gualberti G (1999) The DNA binding site of the Dof protein NtBBF1 is essential for tissue-specific and auxin-regulated expression of the rolB oncogene in plants. *The Plant Cell* 11(3): 323-333.

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

Cooper B, Clarke JD, Budworth P, Kreps J, Hutchison D, Park S et al. (2003) A network of rice genes associated with stress response and seed development. *Proceedings of the National Academy of Sciences of the United States of America* 100(8): 4945-50.

Fornara F, Panigrahi KC, Gissot L, Sauerbrunn N, Rühl, M, Jarillo J, et al. (2009) Arabidopsis DOF transcription factors act redundantly to reduce CONSTANS expression and are essential for a photoperiodic flowering response. *Developmental cell* 17(1): 75-86.

Gualberti G, Papi M, Bellucci L, Ricci I, Bouchez D, Camilleri C, et al. (2002) Mutations in the Dof zinc finger genes DAG2 and DAG1 influence with opposite effects the germination of Arabidopsis seeds. *Plant Cell* 14(6): 1253-1263.

Jahns MT (2008) Avaliação da expressão gênica diferencial entre folhas e tecidos vasculares de *Eucalyptus grandis*. Tese de Mestrado. Universidade Federal do Rio Grande do Sul. Porto Alegre.

Kisu Y, Ono T, Shimofurutani N, Suzuki M, Esaka M (1998). Characterization and expression of a new class of zinc finger protein that binds to silencer region of ascorbate oxidase gene. *Plant & cell physiology* 39(10): 1054-64.

Konishi M, Yanagisawa S (2007). Sequential activation of two Dof transcription factor gene promoters during vascular development in Arabidopsis thaliana. *Plant physiology and biochemistry : PPB / Société française de physiologie végétale* 45(8): 623-9.

Kumar R, Taware R, Gaur VS, Guru SK, Kumar A (2009) Influence of nitrogen on the expression of TaDof1 transcription factor in wheat and its relationship with photo synthetic and ammonium assimilating efficiency. *Molecular biology reports* 36(8): 2209-20.

Li D, Yang C, Li X, Gan Q, Zhao X, Zhu L, et al. (2009) Functional characterization of rice OsDof12. *Planta* 229(6): 1159-69.

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

Medeiros AOS (2009) Avaliação da expressão gênica diferencial entre xilemas contrastantes para a qualidade da madeira de *Eucalyptus grandis* e *Eucalyptus globulus* e estudos funcionais preliminares de genes candidatos. Tese de Mestrado. Universidade Estadual de Santa Cruz. Ilhéus.

Moreno-Risueno MA, Martínez M, Vicente-Carbajosa J, Carbonero P (2007) The family of DOF transcription factors: from green unicellular algae to vascular plants. *Molecular genetics and genomics* : MGG 277(4): 379-90.

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

Umemura Y, Ishiduka T, Yamamoto R, Esaka M (2004) The Dof domain, a zinc finger DNA-binding domain conserved only in higher plants, truly functions as a Cys2/Cys2 Zn finger domain. *The Plant Journal* 37(5): 741-749.

Vicente-Carbajosa J, Moose SP, Parsons RL, & Schmidt RJ (1997) A maize zinc-finger protein binds the prolamins box in zein gene promoters and interacts with the basic leucine zipper transcriptional activator Opaque2. *Proceedings of the National Academy of Sciences of the United States of America*, 94(14): 7685-90.

Yanagisawa S (2002) The Dof family of plant transcription factors *Trends in Plant Science* 7(12): 555-560.

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

Yanagisawa S, Schmidt RJ (1999) Diversity and similarity among recognition sequences of Dof transcription factors. *The Plant Journal* 17(2): 209-214.

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

Zaha A, Ferreira H, Passaglia L (2003) *Biologia Molecular Básica* 3ª Edição. Editora Mercado Aberto. 424 Páginas.

Manuscript to be submitted to “Plant Physiology” (Impact Factor 6.110)

Journal Research Area: Genome Analysis

**COMPARATIVE GENOMICS OF THE *DOF* TYPE GENE FAMILY IN  
*EUCALYPTUS GRANDIS*<sup>1</sup>**

Gabriel Silveira D'Almeida<sup>2</sup>, Michèle Claire Bretton<sup>2</sup>, Giancarlo Pasquali<sup>2</sup>

<sup>2</sup>Biotechnology Center, Federal University of Rio Grande do Sul, Porto Alegre,  
RS, Brazil.

## **FOOTNOTES**

<sup>1</sup>This work was supported by CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Ministry of Education) and CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico, Ministry of Science and Technology).



## **ABSTRACT**

Dof proteins are a family of plant exclusive transcription factors (TFs) that contain a highly conserved zinc finger-like DNA binding domain. These TFs have been found to control the transcription of genes related to plant specific phenomena such as endosperm development, flowering and seed germination. In this work we have identified 18 putative *Eucalyptus grandis* Dof genes both in the available transcriptome and genome. Dof gene sequences were used to perform a phylogenetic analysis with the Dof sequences from *Arabidopsis thaliana* and *Populus trichocarpa*. The amino acid sequences from all 18 Dof sequences matched perfectly with the description of the Dof domain, including the four cysteine residues critical for zinc coordination. Two rooted trees built with the maximum likelihood method and supported by common conserved motifs outside the Dof domain classified the sequences into 10 Major Clusters of Orthologous Genes (MCOGs), allowing us to identify orthologs and paralogs among the three species.

## INTRODUCTION

The Dof family of proteins comprehends plant exclusive transcription factors, characterized by the presence of a highly conserved zinc finger-like DNA binding domain (Yanagisawa, 2002a). These proteins seem to control the transcription of genes related to plant specific phenomena such as endosperm development (Vicente-Carbajosa, Moose, Parsons, & Schmidt, 1997), carbohydrate metabolism (Yanagisawa & Sheen, 1998), plant specific defense mechanisms (Baumann, De Paolis, Costantino, & Gualberti, 1999), seed germination (Papi et al., 2000), flowering (Fornara et al., 2009; Li et al., 2009), and vascular development (Konishi & Yanagisawa, 2007).

The first *Dof* gene (MNB1/ZmDOF1) was isolated from corn, and is related to another *Dof* gene from this species (ZmDof2) in the regulation of the gene that encodes the C4 phosphoenolpyruvate carboxylase (Yanagisawa & Sheen, 1998). Since then, Dof transcription factors have been identified in several plant species, including wheat, barley, rice, tobacco, *Arabidopsis*, pumpkin, potato, pea and papaya (Yanagisawa, 2004). These genes appear to have originated from a common ancestor that kept a single copy in the unicellular algae *Chlamydomonas reinhardtii*, and afterwards, expanded to different taxonomic groups, increasing in number by means of duplication events (Moreno-Risueno, Martínez, Vicente-Carbajosa, & Carbonero, 2007).

The DNA binding Dof domain is characterized by a highly conserved amino acid sequence, with two pairs of cystein residues that create the zinc finger-like motif by interacting with a zinc (Zn) atom (Umemura, Ishiduka, Yamamoto, & Esaka, 2004; Yanagisawa & Schmidt, 1999). Experiments with metal chelators were able to prevent the binding of these proteins to DNA, and the addition of Zn was able to restore this capacity afterwards. The substitution of the cysteine residues with serine and alanine also inhibited the binding (Yanagisawa, 2004).

The Dof domain is also characterized by its capacity to perform protein-protein interactions, being classified as a bifunctional domain (Yanagisawa, 2004). Interactions between proteins mediated by the Dof domain were already described in *Arabidopsis* and corn (Yanagisawa, 2004). In both cases, the Dof proteins were capable of interacting with stress response proteins containing a leucin zipper

domain. Since the Dof domain is highly conserved, it is presumed that all Dof proteins should be able to perform protein-protein interactions *in vivo*.

In order to better understand the importance of *Dof* genes and proteins in *Eucalyptus* trees, especially in what concerns wood biogenesis and defense mechanisms, we started an effort to characterize this gene/protein family in *Eucalyptus grandis*. In this study, we describe the identification of 18 *E. grandis* *Dof* genes based on both transcriptome and genome. We performed a phylogenetic analysis including *Dof* gene sequences from *E. grandis*, *Arabidopsis thaliana* and *Populus trichocarpa*.

## RESULTS

### Obtaining *Dof* Sequences from Databases

In order to identify *Dof* nucleotide sequences within the *E. grandis* genome, we first obtained the complete nucleotide and amino acid *Dof* sequences from nine different species from the online databases The Arabidopsis Information Resource and Plant Transcription Database. Species included were *Arabidopsis thaliana*, *Arabidopsis lyrata*, *Carica papaya*, *Populus trichocarpa*, *Vitis vinifera*, *Sorghum bicolor*, *Chlamydomonas reinhardtii*, *Oryza sativa indica* and *Zea mays*. Since *Dof* sequences show little identity, we performed individual local BLAST searches on the *E. grandis* genome (EucaGen Genome Browser) using the putative amino acid *Dof* domain sequence from each of the complete amino acid sequences. The resulting scaffolds were confirmed by individual BLAST searches through three different databases. With this method, we were able to obtain 20 putative *E. grandis* *Dof* sequences. The sequences were named EucaDof1 to 20, according to confirmation order.

From these initial *E. grandis* *Dof* genes identified within the EucaGen Genome Database, EucaDof8 was the only sequence found complete. EucaDof1, 2, 3, 9, 11 and 13 to 19 presented incomplete coding regions but with an initial methionine start, whilst EucaDof 4 to 7, 10, 12 and 20 were incomplete, without the initial starting methionine. The complete sequences were obtained by individual searches on online databases. Expression as mRNA was confirmed by individual searches on the Genolyptus EST Database.

### Alignment and Phylogenetic Analysis

An initial alignment with the 20 putative *Dof* gene sequences showed that sequences EucaDof13 and 14 were the same, and that EucaDof20 did not have a complete *Dof* domain (data not shown). Sequences EucaDof13 and 20 were then excluded from further analysis and the confirmed sequences were renamed as EucaDof1 to 18. Individual BLAST searches on the NCBI protein database revealed that the 18 sequences present high identity with *Dof* proteins from *A. thaliana*, *P. Trichocarpa* and *Vitis vinifera*. (Table I).

A new alignment was made with the 18 putative Dof domain sequences, as shown in Figure 1, in order to evaluate the evolutionary relationship among them. The sequences aligned perfectly within the Dof domain and, as expected, showed little identity outside this motif.

In order to further analyze the relationship among Dof protein sequences, two rooted trees were built by the maximum likelihood method. The first tree was built with the 18 putative Dof amino acid sequences (Figure 2), whilst the second was built with the 18 *E. grandis* protein sequences plus the sequences from *A. thaliana* and *P. trichocarpa* (Figure 3). The first tree presented possible paralogs (EucaDof15 and 6, and EucaDof12 and 1), whilst with the second tree was possible to identify 10 Major Clusters of Orthologous Genes (MCOGs), that were colored and numbered from I to X. Nine of these clusters contained putative *E. grandis* Dof genes. When possible, EucaDof genes were further classified in pairs or smaller orthologous subgroups (marked in red on Figure 3). From the six determined orthologous pairs, three were combined with *A. thaliana* genes (EucaDof17/AT5G60850.1; EucaDof16/AT3G21270.1; EucaDof5/AT5G62940.1), and three with *P. trichocarpa* genes (EucaDof10/n824514; EucaDof4/n209054; EucaDof3/n820338). From the three subgroups, the first was composed exclusively with *E. grandis* genes (EucaDof13/ EucaDof12/ EucaDof1), the second contained genes from the three species (EucaDof18/AT3G50410.1/n286802/n563045/n760414/AT5G66940.1), and the third contained sequences from *E. grandis* and *P. trichocarpa* (EucaDof15/EucaDof6/EucaDof14/n577658/n556324/n803398).

The search for conserved motifs was performed on the *A. thaliana*, *P. trichocarpa* and *E. grandis* putative Dof sequences with the MEME Suit software. Fifty motifs of up to 100 residues were extracted, if present, in at least two different sequences. This analysis revealed the existence of homologous motifs conserved among Dof protein sequences, that were different from the Dof domain (Table II). The motifs were numbered from 1 to 50. Surprisingly, the Dof domain (Motif 1), was present only in 93 of the 95 sequences used, being absent in two *P. trichocarpa* genes (248810 and 256257). The Dof domain, however, was found in all *Populus* proteins when employing the MUSCLE alignment (data not shown).

*A. thaliana* sequences held the largest number of different conserved motifs (43), missing only motifs 3, 4, 5, 24, 32, 37 and 39. Motifs 13 and 40 were found twice in sequences AT4G00940.1 and AT4G55370.1 respectively. Motifs 31, 38, 44

and 50 were found exclusively in Dof protein sequences from *Arabidopsis*. *P. trichocarpa* Dof protein sequences presented 39 different motifs, missing the ones numbered 20, 21, 24, 28, 31, 36, 38, 39, 42, 44 and 50. Motifs 3 and 32 were found exclusively in Dof protein sequences from *Populus*. *E. grandis* Dof sequences presented 42 different motifs, missing the ones numbered 3, 22, 31, 32, 38, 44, 48 and 50. Motif numbers 2, 19, 25 and 39 were found twice in sequences EucaDof17, 4, 10 and 1, respectively. Motif numbers 24 and 39 were found exclusively in *E. grandis* Dof proteins. From the 50 extracted motifs, 31 were shared by the three species analyzed, and 8 were found exclusively in one of them.

## DISCUSSION

We have identified 18 putative *Dof* genes within the *E. grandis* genome. Our results, based on sequence homology, indicate that these sequences are functional *Dof* genes (functionality can be assumed by the presence of the extremely conserved *Dof* domain, as previously described by Yanagisawa et al. (2002)), with motifs in common with *A. thaliana* and *P. trichocarpa* *Dof* genes.

The *P. trichocarpa*, *A. thaliana* and *V. vinifera* putative *Dof* sequences found by individual BLAST searches on NCBI, as well as the presence of the 18 *E. grandis* sequences on the Genolyptus EST database, further support this assumption.

The fact that only one coding sequence (EucaDof8) was found complete was possibly due to the fact that *Dof* proteins share little structural similarity besides the *Dof* domain itself, and that the search was performed with the putative *Dof* domain from the other species. Similar results were obtained by Yanagisawa et al. (2004), Lijavetzky et al. (2003) and Costenaro da Silva et al. (2010) when searching for *Dof* sequences in *Arabidopsis*, *Rice* and *Vitis*.

Analysis of the first rooted tree, build with the 18 putative *Dof* aminoacid sequences using the maximum likelihood method suggests that EucaDof17 is the most conserved sequence in relation to the original gene (root). EucaDof 15 and 6, and 12 and 1 were the most structurally similar, and possibly paralogs. Since paralogs usually display different functions within organisms, no functional relationship can be presumed from this tree (Lijavetzky, Carbonero, & Vicente-Carbajosa, 2003).

The second rooted tree built with the maximum likelihood method presented 10 MCOGs, six pairs and three subgroups of orthologous genes containing the *E. grandis* sequences. The orthologous subgroups depicted here may contain important information about the function of the recent discovered *E. grandis* *Dof* genes, since orthologous genes usually display similar functions (Lijavetzky et al. 2003).

The trees showed several clusters with low bootstrapping support. However, most of the groups and subgroups described here can be supported by the presence of common protein motifs outside the *Dof* domain, detected in the MEME analysis, as previously stated by Lijavetzky et al. (2003).

As far as our knowledge goes, this is the first work reporting the profile of *Dof* genes in the genome and transcriptome of an *Eucalyptus* species. The sequences

expression as mRNA was confirmed by the Genolyptus EST Database. We intend to further characterize the 18 *Dof* genes, describing their profiles of expression via quantitative RT-PCR. Heterologous expression in *Escherichia coli* and biochemical characterization of the recombinant proteins is also an intended objective. Finally, overexpression or repression of *Dof* gene expression in transgenic plants will allow us to better functionally characterize the described *Dof* sequences.



## MATERIALS AND METHODS

### Database Search

Two hundred and sixty Dof amino acid sequences from nine different species were obtained from online databases. These species include *Arabidopsis thaliana* (36 sequences), *Arabidopsis lyrata* (34 sequences), *Carica papaya* (18 sequences), *Populus trichocarpa* (41 sequences), *Vitis vinifera* (26 sequences), *Sorghum bicolor* (30 sequences), *Chlamydomonas reinhardtii* (1 sequence), *Oryza sativa indica* (17 sequences) and *Zea mays* (57 sequences). The *Arabidopsis thaliana* sequences were obtained from The Arabidopsis Information Resource (<http://www.arabidopsis.org/>), whilst the remaining sequences were obtained from the Plant Transcription Factor Database (<http://planttfdb.cbi.pku.edu.cn/>). In order to avoid low score scaffolds and pseudo genes in the local BLAST search, the sequences were then used to perform independent searches on the Pfam database (<http://pfam.sanger.ac.uk/>), from where the putative Dof domain sequences were obtained (accession number PF02701).

### Local BLAST Search and Sequence Confirmation

The putative Dof domain sequences were used on independent local BLAST searches through the *E. grandis* genome. High score scaffolds resulting from this procedure were used on independent BLAST searches in three different databases, including NCBI (<http://blast.ncbi.nlm.nih.gov/>), LGE (<http://www.lge.ibi.unicamp.br/eucalyptus/>) and SisGen (<http://genoma.embrapa.br/>). This resulted in 19 incomplete and 1 complete putative Dof sequences. In order to obtain the complete sequences we performed individual BLAST searches through the Populus EST Database (<http://www.populus.db.umu.se/blast.php>), then, tBLASTx searches through NCBI database, with the resulting scaffolds used to perform individual BLAST searches through the *E. grandis* genome database (EucaGen Database Browser at <http://eucalyptusdb.bi.up.ac.za/>).

### Alignment and Phylogenetic Analysis

Alignment of the 20, and later 18 putative amino acid Dof sequences, as well as the *Dof* gene sequences from other species, was conducted with the MUSCLE software (BLOSUM62 Algorithm), with the input files in FASTA format, and the results

where visualized with the GeneDoc software. The two rooted trees were build with Phylogeny.fr online software (<http://www.phylogeny.fr/>), using the "one-click" maximum likelihood method, and visualized with the online software Interactive Tree of Life (<http://itol.embl.de/>). The first tree was build using the 18 putative Dof amino acid sequences (Figure 1), whilst the second was build with the 18 *E. grandis* sequences plus the sequences from *A. thaliana* and *P. trichocarpa* (with a total of 95 sequences). All sequences were inserted in FASTA format. A search for conserved motifs was performed using MEME Suit Software. The Dof amino acid sequences from *E. grandis*, *A. thaliana* and *P. trichocarpa* were inputted in FASTA format, and MEME was set to search for a maximum of 50 conserved motifs from 10 to 100 residues, that appeared at least in two sequences. The results were exported and analyzed in HTML format.

## LITERATURE CITED

Baumann K, De Paolis A, Costantino P, Gualberti G (1999) The DNA binding site of the Dof protein NtBBF1 is essential for tissue-specific and auxin-regulated expression of the rolB oncogene in plants. *The Plant Cell* 11(3): 323-333.

Costenaro da Silva et al. (Submitted) Genome-wide comparative analysis of the Dof-type gene family in grapevine. *Plant Physiology*.

Fornara F, Panigrahi KC, Gissot L, Sauerbrunn N, Rühl M, Jarillo J, et al. (2009) Arabidopsis DOF transcription factors act redundantly to reduce CONSTANS expression and are essential for a photoperiodic flowering response. *Developmental cell* 17(1): 75-86.

Konishi M, Yanagisawa S (2007) Sequential activation of two Dof transcription factor gene promoters during vascular development in *Arabidopsis thaliana*. *Plant physiology and biochemistry : PPB / Société française de physiologie végétale* 45(8): 623-9.

Li D, Yang C, Li X, Gan Q, Zhao X, Zhu L, et al. (2009) Functional characterization of rice OsDof12. *Planta* 229(6): 1159-69.

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

Moreno-Risueno MA, Martínez M, Vicente-Carbajosa J, Carbonero P (2007) The family of DOF transcription factors: from green unicellular algae to vascular plants. *Molecular genetics and genomics : MGG* 277(4): 379-90.

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

Umemura Y, Ishiduka T, Yamamoto R, Esaka M (2004) The Dof domain, a zinc finger DNA-binding domain conserved only in higher plants, truly functions as a Cys2/Cys2 Zn finger domain. *The Plant Journal* 37(5): 741-749.

Vicente-Carbajosa J, Moose SP, Parsons RL, Schmidt RJ (1997) A maize zinc-finger protein binds the prolamin box in zein gene promoters and interacts with the basic leucine zipper transcriptional activator Opaque2. *Proceedings of the National Academy of Sciences of the United States of America* 94(14): 7685-90.

Yanagisawa S (2002) The Dof family of plant transcription factors. *Trends in Plant Science* 7(12): 555-560.

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

Yanagisawa S, Schmidt RJ (1999) Diversity and similarity among recognition sequences of Dof transcription factors. *The Plant Journal* 17(2): 209-214.

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

## FIGURE LEGENDS

**Figure 1.** Multiple sequence alignment of 18 predicted Dof domain sequences from the *E. grandis* genome. The zinc finger location is marked from the first to the second cysteine pair. Conserved amino acids are indicated by a black column, whilst semi-conserved amino acids are indicated by gray columns.

**Figure 2.** Phylogenetic tree of the putative *Dof* gene family members from *E. grandis*. The confidence probability (multiplied by 100), that the interior branch length is greater than zero, as estimated using the bootstrap test, is shown under the branch. Bootstrap values <20 are not shown. Phylogenetic analysis were conducted with the IToL online software.

**Figure 3.** Phylogenetic tree of the putative Dof gene family members from *E. grandis*, *A. thaliana* and *P. trichocarpa*. The resulting MCOGs are numbered from I to X. The confidence probability (multiplied by 100), that the interior branch length is greater than zero, as estimated using the bootstrap test, is shown under the branch. Bootstraps values <20 are not shown. Phylogenetic analyses were conducted with the IToL online software.

**Figure 4.** MEME alignment performed with the complete amino acid sequences from *A. thaliana*, *P. trichocarpa* and *E. grandis*. The extracted motifs are presented as colored blocks numbered from 1 (Dof domain) to 50. The sequences are presented in the same order they were inputted in the FASTA file. Regular expression for the defined motifs are presented in Table II.

## Tables

**Table I.** *Eucalyptus grandis* Dof genes.

Name <sup>a</sup>	Protein Length <sup>b</sup>	Genolyptus EST Access <sup>c</sup>	NCBI BLASTp <sup>d</sup>	Accession <sup>e</sup>	Max Score <sup>f</sup>	e-value <sup>g</sup>
EucaDof1	270	GR-PU-003&&MG-047&&D04.ab1&&113390349287137868559744306811	unnamed protein product [V. vinifera]	CBI1625 5.3	163	2e-38
EucaDof2	307	CL4605Contig1&&113391157472646611893099409932	F-box family protein [P. trichocarpa]	XP_0023 03678.1	335	4e-90
EucaDof3	342	CL9546Contig1&&113391402058613338495854679007	unnamed protein product [Vitis vinifera]	CBI3056 9.3	290	2e-76
EucaDof4	227	CL6277Contig1&&113390783317859448543883783664	f-box family protein [P. trichocarpa]	XP_0023 06453.1	216	1e-54
EucaDof5	364	GR-ML-002&&BA-081&&F03.esd&&113390311826916988382838732809	hypothetical protein ARALYDRAFT_919576 [A. lyrata subsp. lyrata]	XP_0028 66523.1	512	3e-143
EucaDof6	437	GR-ML-001&&BA-071&&D03.esd&&113390308672544817733142713800	predicted protein [P. trichocarpa]	XP_0023 23863.1	555	6e-156
EucaDof7	307	SP-RX-003&&GO-009&&F08.ab1&&113390597223478069102165295382	zinc finger protein, putative [R. communis]	XP_0025 24061.1	627	5e-178
EucaDof8	265	CL1894Contig2&&113390733589899311085534039133	predicted protein [P. trichocarpa]	XP_0023 20208.1	268	4e-70
EucaDof9	271	GR-TS-001&&AF-026&&EUGR-TS-001-026-F06-AF.R.ab1&&11339040996	predicted protein [P. trichocarpa]	XP_0023 20208.1	383	1e-104
EucaDof10	316	CL1894Contig1&&113390733538894951745250783805	PREDICTED: hypothetical protein [V. vinifera]	XP_0022 75610.1	405	4e-111
EucaDof11	366	SP-FX-001&&CN-054&&CN-EUSP-FX-001-054_B11_g_089.ab1&&113390	PREDICTED: hypothetical protein [V. vinifera]	XP_0022 74028.1	268	1e-69
EucaDof12	195	GR-XY-022&&BA-038&&EUGR-XY-022-038-B10-BA.R.ab1&&11339049449	zinc finger protein, putative [R. communis]	XP_0025 32252.1	268	3e-70
EucaDof13	285	GR-XY-022&&BA-006&&F04.esd&&113390487697461356810925433110	elicitor-responsive Dof protein ERDP [P. sativum]	BAA856 55.1	518	3e-145
EucaDof14	458	CL9660Contig1&&113390999674049487680843255102	PREDICTED: hypothetical protein [V. vinifera]	XP_0022 83706.1	681	0.0
EucaDof15	437	GR-TS-001&&AF-025&&EUGR-TS-001-025-C06-	predicted protein [P. trichocarpa]	XP_0023 23863.1	555	6e-156

---

		AF.R.ab1&&11339040940353432209 0119702709				
EucaDof16	181	CL2513Contig1&&11339113029526 5360799491102017	ADOF1; DNA binding / transcription factor [A. thaliana]	NP_1755 81.1	356	6e-97
EucaDof17	219	GR-SE-001&&GO- 163&&G05.ab1&&11339039450338 3835335579730371	zinc finger protein OBP4 - like [A. thaliana]	BAE992 82.1	189	2e-46
EucaDof18	239	CL3369Contig1&&11339259776701 7706032324091240	predicted protein [P. trichocarpa]	XP_0023 10762.1	370	8e-101

---

<sup>a</sup>According to confirmation order, excluding EcaDof13 and 20.

<sup>b</sup>Number of amino acids residues contained in the protein.

<sup>c</sup>Genolytus EST Database Accession.

<sup>d</sup>Performed on the non-redundant protein sequence database (nr).

<sup>e</sup>NCBI Protein Database Accession Number.

<sup>f</sup>NCBI highest alignment score of a set of aligned segments from the *Dof* gene and the database protein.

<sup>g</sup>Number of alignments expected by chance with the particular score.

**Table II.** Conserved motifs extracted from *E. grandis*, *A. thaliana* and *P. trichocarpa* *Dof* genes employed in this study.

Motif Number <sup>a</sup>	Protein Length <sup>b</sup>	Regular Expression of the Extracted Motif <sup>c</sup>
1	43	[KP]CPRC[DN]STNTKFCYNNY[SN]L[ST]QPR[HY]FCK[TS]CRRYWTKGG[TAS]LRN[VP]
2	18	VGGG[CRS]RKNK[RK]S[SK]SSSSSS
3	36	CPRC[DG]SSNTKFCYNNY[SN]LTQPR[HY]FCK[GT]CRRYWTKG
4	80	MSEPKDQAFKLFGKTIQVPEISVTTATTTDDDDDDSDQDRPSCANSSLEDTNITDDY NNNDKRDHGEEDTETDDKDSV
5	70	[KE][TKV][TS][IDL][ET][NER][QK][EQ][DE][GD][AV][SR][PS][VS][AT][AST][KE]E[SF P][ST][ND][LP]DA[TN]SG[TV][SN]ENP[KR]TPS[VAI][DE]KE[ST][TAE][TAG][LP][KRQ ][TC]S[KNR][STN]EEE[QD]S[DE]TS[NI]S[QP]EK[TV]LKKPKDKI
6	19	TLRIDDP[SGN]EA[AS]KS[SP][IL]W[AT]TL
7	42	NGV[HT][HN]P[PS][LS][KG][NP]NGTVLTFGSD[AS]PL[CH][ED]S[MV]ASVL[NHL]L[A S][DEK]K[TA][MQR][HN][NS]S[RT]
8	25	[KGN][HN]V[AS][EA]T[TS][PS]VLQANPAALSRSLNF[HQ]E
9	26	MDTA[QK]WPQE[IF]VVKP[ILM][EN][ED]IV[TV][NS]TC[PLST][KA][PAQ]
10	14	[KR]DP[AG]IKLFGK[TK]I[PT][LF]
11	30	[AT]FCP[PS]GFP[VM][PS]FYPA[PA]AYWGCTVP[GS][AP]WN[VI]P[WP]
12	24	G[LYS][GE]S[LRG][YHQV]G[VI][QH]E[NST][NSG][GA]RLFPF[EG][DE]L[KM][QK]
13	16	[HY]EG[QN]DL[NS]LAF[PA]S[LT]QK[QS]
14	19	P[PN][QGS][LV]PC[FY][PS][GV][PVA][PS]W[PV][YF][PAT]WN[SP]A
15	10	ER[KR][ALP]RPQ[KEH][DEQ][QL]
16	18	[GQ]QG[SDG]S[AS]GYWNGM[LF]GGGSW
17	50	[FS]HKP[ED][AG][VL][IR][MIV]P[SV][PS][CY]G[SG][GV][ER][NS]G[DV][DE][HT][MS] NGSS[AV]T[AV][AS]NS[NIS][DL][EV][AE][GS]K[NS][GMV][ST]KE[IS][AV]M[PQ]N[C Y]
18	14	M[AV][DE]RAR[LQ]A[KN][IV]P[LPQ]P[EA]
19	19	MVFSS[VI]P[ASV]YLDP[PSA][NS][WN][QW]QQ[QP]
20	80	LNHGVLNNNNNNGGFNGISTGGNGGGGLMDISTCQRLMLSNDHHHYNHQEDH QRVATIMDVKPNPKLLSLDWQQDQ[CY]
21	70	HFGGMMGSYSTPEHGNVGFLESKYGGLLSQSPRPIDFLDSKFDLMGVNNDNLVMVNH GSNGDHHHHHHHH
22	70	FQP[LS]NVYNY[ST][EG]SMEDSTITT[IT]MP[PS]TSSTI[AT][HQ][PS][LW]QVPNTSSGM DMTNYWNWDDIENYVS[AT]DLNV[PT]WDDSEI
23	13	[ST][PS]TLGKHSRDEN
24	70	LIGVAVCQVRGTFGLPSQLRRTTAPRAPHLILRPWGNTRGMRKSGNQTTPLKRSQRR LILRDPCGFQRH
25	19	[SY]S[SG]GFPLQE[LF][KR]P[TS]LSF[SP][LA]D
26	34	[ST]RV[ST]Q[LT]APVK[MT]E[DEG]NQGLNLS[KR]P[VY][LS]G[IL][PT][GS][NP][DGN] [QN][YQ]Y[NW]
27	35	[TP][GS][GP][AG][SG][GF][LQ]A[LD][GMN]G[FL][GQ][LI]G[SIL]G[SP]G[IM]E[DS][VS] [GNS][FI][GA][LS][GA][RG]G[LV]W[PQ]F[PS]
28	64	LKNQTVAEKPDHGGSGSEEKEERVSGQEMNPTRMLYGLPVGDPNGASFSSLLASNMQ MGGLVYE
29	26	[LE]L[LK][RS]SG[FI]ASRGLN[TS]F[IM]PTPM[PS]DSN[TS]L
30	10	[EL][KT][TA][LE]K[KR]P[DT]KI
31	61	V[AG]AVGN[LR]F[GS]SL[SY][DH]IHGGMVTN[LV]HPT[QR]T[FV]RPNHRLAFH[DN]G



		FEQDYDVGSDNLLVNQQVGGY
32	100	NM[IT]N[RW]VNPQ[LQ][AP]IQAQQQKQNLPLDLVGLG[DG]QK[DN][LS][ES]ILYQAMIN PPSSVLQQNSISCNN[FL]DTKSFN[TV]NNNGVLLGSS[LV]S[IL]PQ[IN]QG[LS]LFP[CY][ PS]SSFDTHPS[LS]ICT
33	19	[KQ][SKP][DN][SP][IA][GNS][GK][GR][GT][LI]FK[AGN]F[DQE][ST]K[ASV]E
34	49	MQDIHDFSMNGVGGGGGGGRFFGGGIGGGGGGDRRMRAHQNNILNHHQ
35	26	ID[GN]LPLPATPV[L]VPLTA[NS]QG[LF]SVHFG
36	50	MDSDWLQESESSGGSM LDSSTNSPSAADILAACSTRPQASAVAVAAAALM
37	36	MP[AT]ELSSSE[AK][AT]RR[AP][QP][QL]ST[AT][AP]T[LS][TN]KPGGAPPQEQEHL
38	19	M[DN]N[LF]NV[FV][AT]NEDN[QE][VM]N[DG][VLM][KP][PR]
39	50	FDDQHDFSILGNPVSSHCDILGNPSSSSTATTTPAFLDALRSGFLDTH
40	26	[QP][QT][NQ][SP]Q[FL]P[FI][LM][PT][PST]L[HQ][NH]L[TG][QD][FLY][GN][SG][GI][N G][IL][GN]L[NAD]
41	13	S[SKP][KQ]LPDL[N]PP[IPS][LV][FLS]
42	12	[RK]V[FL][WG]GFPWQ[MI][NT][GM]
43	26	[WD][GLQ][FGP][GE][SM][GD][LG][GQ][GLN]G[FS]V[GR][GR][SD][CD][FD][ST]W[PT ][DG]L[AI][IM][SNT][TPR]
44	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
45	18	M[QD][DH][LHQ][THQ][SHY][AH][AH][AHV][YDH][YMQ][HSY]Q[SFH][MLQ][MQ]M [TP]
46	13	IDLA[LA]V[YF]A[NKQ][FY][LV][NT][PD]
47	10	[QR]QQQQ[QN]Q[PL]Q[QF]
48	24	IE[ST]LSC[IL]NQDLHW[KR]LQQQR[LM]AMLF
49	10	[NK][PTA]E[RK][CS][LI][WL][IV]PK
50	23	YHMN[PQ]VDQ[FY][KN]WNQSFNN[AT]MNMNY

<sup>a</sup>Numbers are the same from Figure 4. Motif 1 corresponds to the Dof domain.

<sup>b</sup>Number of amino acid residues contained in the motif.

<sup>c</sup>Letters with observed frequency greater than 0,2 (MEME default parameter). Sequences obtained from the analysis of the 95 Dof amino acid sequences from *A. thaliana*, *P. trichocarpa* and *E. grandis*.

## Original Figure Files

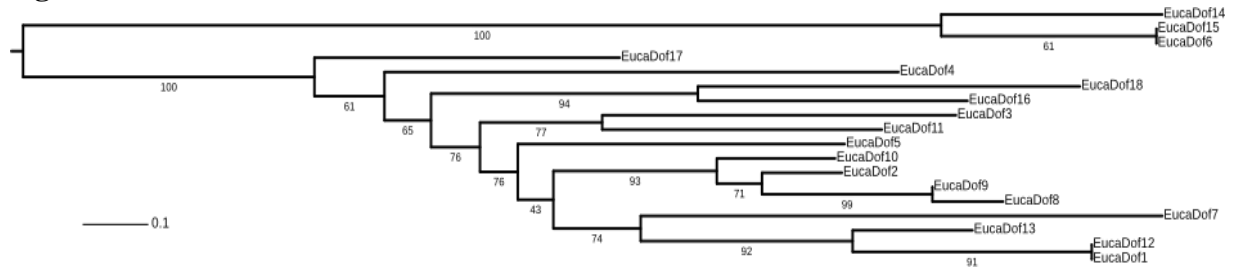
**Figure 1**

```

----- Zinc Finger -----
EucaDof6 : PDKLLECPRCNSMDTKFCYYNNYNNVQPRHFCKNCQRYWTAGGTMRNVPVGAERRKNRNSASHYRH
EucaDof15 : PDKLLECPRCNSMDTKFCYYNNYNNVQPRHFCKNCQRYWTAGGTMRNVPVGAERRKNRNSASHYRH
EucaDof14 : PDKLLECPRCNSMDTKFCYYNNYNNVQPRHFCKNCQRYWTAGGTMRNVPVGAERRKNRNSASHYRH
EucaDof11 : QQQPLECPRCSSNTKFCYYNNYSLSQPRHFCKACKRYWTRGGTLRNVPVGGCRKNRNVKRPASA
EucaDof5 : HHQPQKCPRCDSMNTKFCYYNNYSLSQPRYFCKTCRRYWTRGGTLRNVPVGGCRKAKRPFSSSNSS
EucaDof17 : HHQSTKCPRCNSLNTKFCYYNNYSLSQPRHFCKNCRRYWTRGGVLRNVPVGGCRKAKRPSKTSGLIP
EucaDof4 : NPENLECPRCDSANTKFCYYNNYSLTQPRHFCKTCRRYWTRGGALRNVPVGGCRKNRNSAVSSSA
EucaDof8 : PQEQLECPRCNSNNTKFCYYNNYSLTQPRYFCKTCRRYWTRGGSLRNVPVGGSRKNRNVSTSSSTS
EucaDof9 : PQEQLECPRCNSNNTKFCYYNNYSLTQPRYFCKTCRRYWTRGGSLRNVPVGGSRKNRNVSTSSSTS
EucaDof2 : KEQALECPRCNSNTKFCYYNNYSLTQPRYFCKTCRRYWTRGGSLRNPVGGSRKNRNSSSSSAS
EucaDof10 : KDQALECPRCTSTHTKFCYYNNYSLSQPRYFCKTCRRYWTRGGSLRNVPVGGSRKNRPFSSSSPS
EucaDof7 : VEMSPNCPRCSSNTKFCYYNNYSLTQPRYFCKTCRRYWTRGGSLRNVPVGGCRKNRNVGSLRLSS
EucaDof16 : MTTQLECPRCDSNTKFCYYNNYSLSQPRHFCKSCRRYWTRGGALRNIPVGGTRKTAIRASSKSS
EucaDof3 : PEAALECPRCDSANTKFCYFNYSLSQPRHFCKTCRRYWTRGGALRNVPVGGCRKNRNSKSSSK
EucaDof18 : EQEHLLECPRCDSNTKFCYYNNYSLSQPRHFCKSCRRYWTRGGTLRDI PVGGTRKNARSRRTSAT
EucaDof13 : P-EAHLCPRCBSNTKFCYYNNYSLLQPRYFCKSCRRYWTRGGTLRNVPVGGCRKNRNSSSSSSS
EucaDof1 : PEEAHLCPRCBSNTKFCYYNNYSLLQPRYFCKSCRRYWTRGGTLRNVPVGGCRKNRNSSSSSSS
EucaDof12 : PEEAHLCPRCBSNTKFCYYNNYSLLQPRYFCKSCRRYWTRGGTLRNVPVGGCRKNRNSSSSSSS
1 C PRC S TKFCY5NNY QPR FCK C RYWT GG 6R16P6Gg R4n

```

**Figure 2**



**Figure 3**

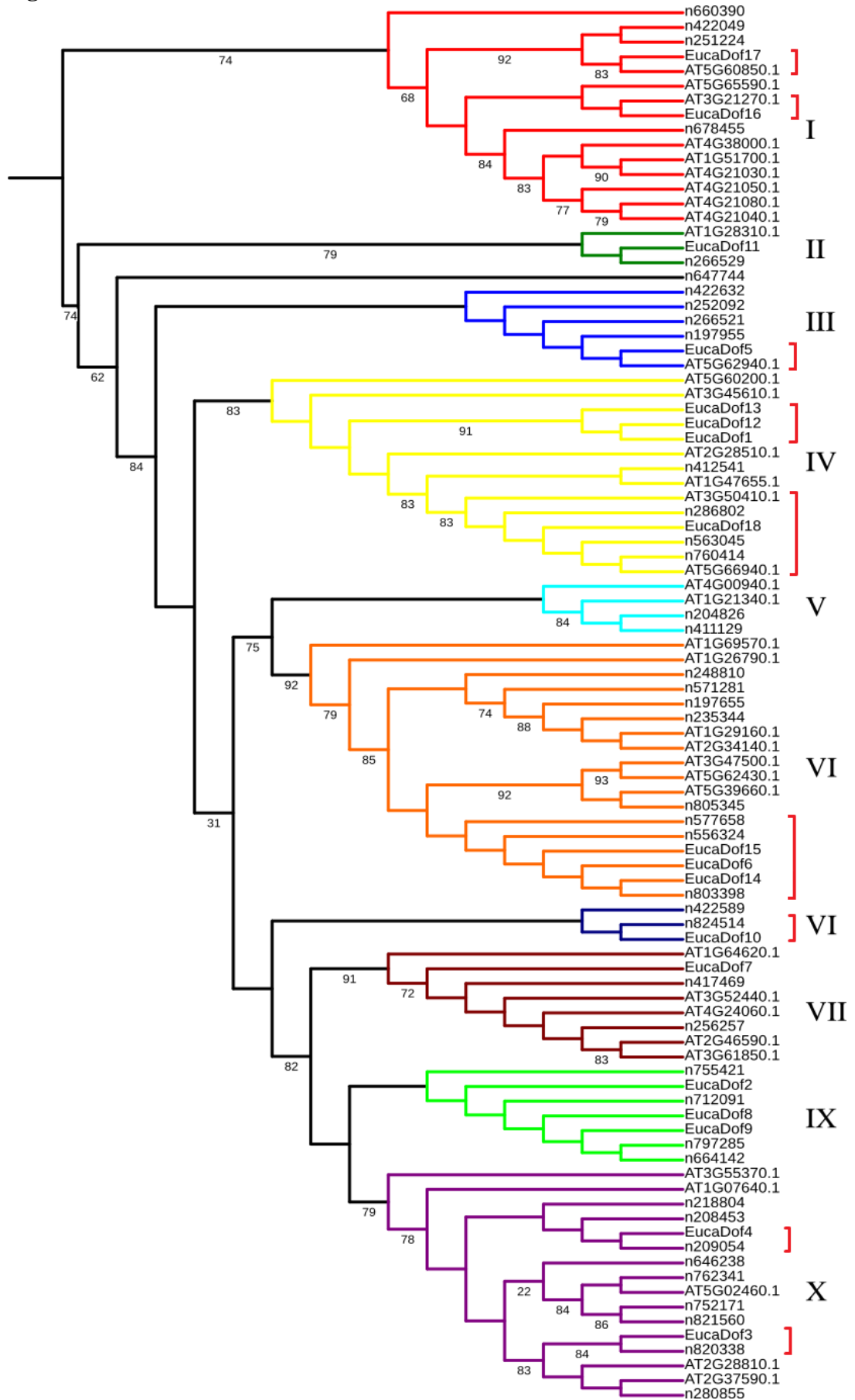
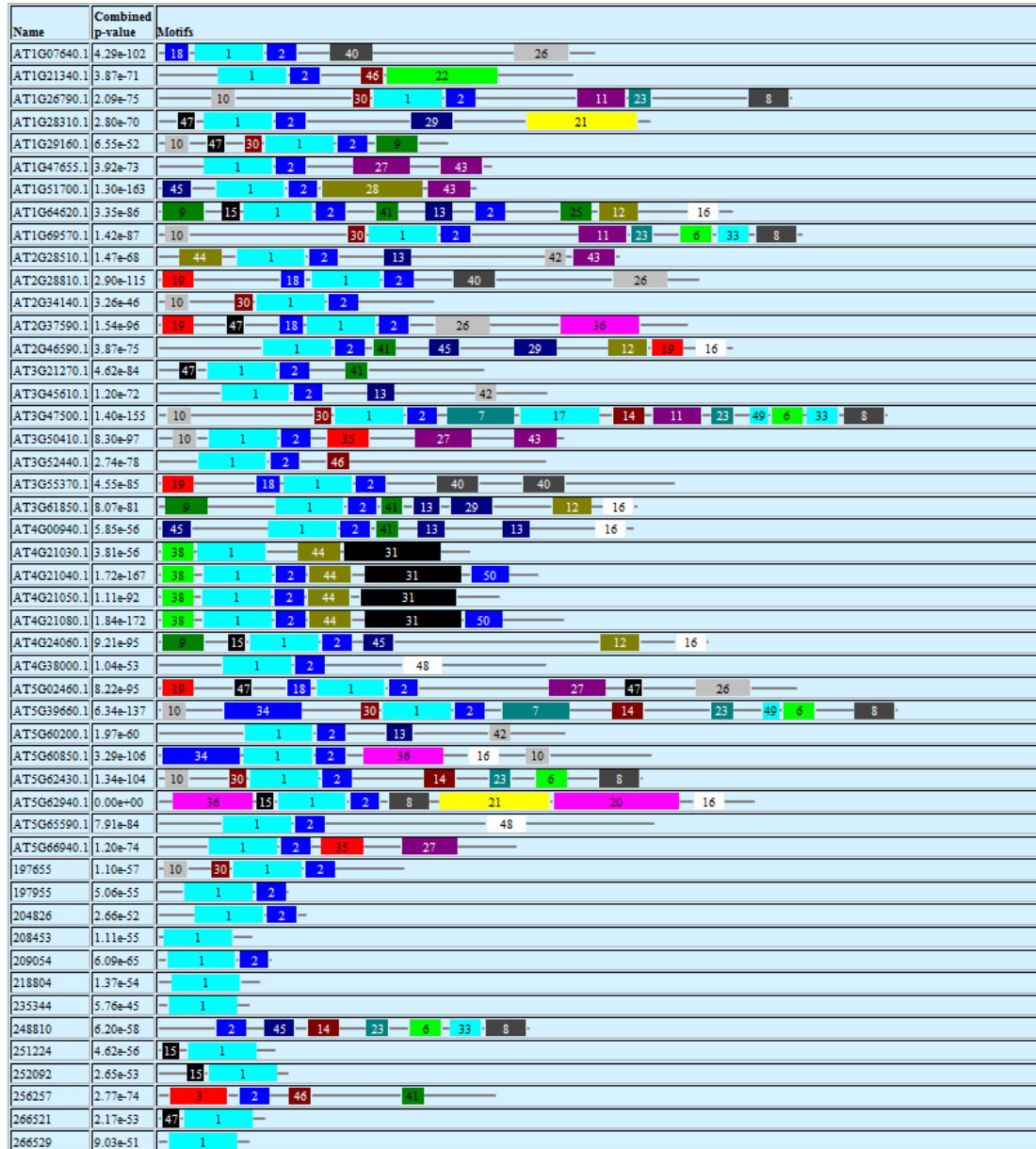
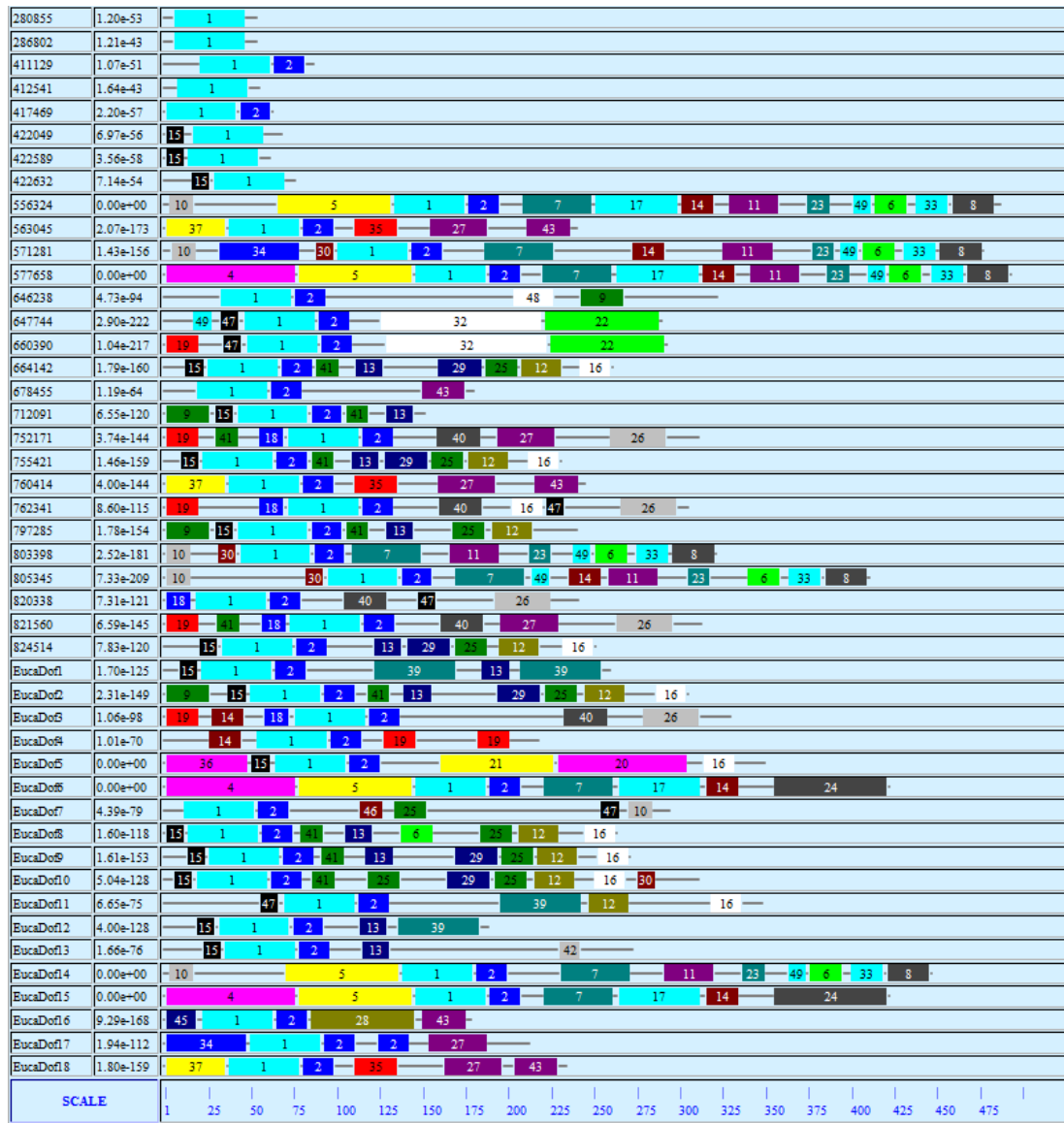


Figure 4





## Supplemental Material

### Supplemental Text I. *Eucalyptus Dof* protein full sequences in FASTA format.

>EucaDof1

MLISPKPQQDQRKPRPQPEEALHCPRCESTNTKFCYNNYSLLQPRYFCKSCRRYWTKGG  
TLRNVVPGGGCRKNRRSSSSSSRRRAQDQGGPISTLINNNDYDTPNPFITLHSLTYDQSS  
DHMSLPFLGLHKFPYQLGPELSDHILSYANRSKNGHGMNGGFLDSSSKRSQDQPLMSS  
NTSPLTNFPPLSYDTNDLSLAFARLQKQNGGQLGFDDQHDFSILGNPVSSHCDILGNPSI  
SSSTSATTPAFLDALRSGFLDTHNNFQNL

>EucaDof2

MDTAQWPQEIVVKPIEDIVTSTCTAAATPKPSSSSVSEKPRPQKEQALNCPRCNSNTNK  
FCYNNYSLTQPRYFCKTCRRYWTDGGLRNIPVGGGSRKNKRSSSSASSSSSFNSSSK  
KLPDLISTPASNPNNKVTLHEGQDLNLAFPNPHHDFKSISELVQVPSLEASKNHHISAN  
SSSAGASMAPPQLSALELLSGITSRGFSFMSMPVHDPGSVYTPGLFALPDFSRPLNFS  
LDGLNGYGSLSHGVTETGRLLPFTEITVEDLKQVVPSTNEFDHQQRNSQEGSAHGYWNG  
MLGGGSW

>EucaDof3

MVFSSMPSYLDPANWQQPHGNSGLNNQLPPAPPPPPPPPHGGAGGNPGGSIRPGSMV  
ERARLANIPMPEAALKPCRCDSANTKFCYFNYSLSQPRHFCKTCRRYWTRGGALRNVPV  
GGGCRNRKRSKSSSKSPITSGSSDQASGGVSSITSRGSSGLAGDILNRATAQIPGLR  
YMSPSSLNNGELGNSGEMGLNYGGIPGSMVGDAGNLSFQLGGNSSLLSMGSLEHQWRLH  
QVTQQFPFMSSLDKHANDLYSMDASTSGVEPSGLAGHHQVRPNLFMNSGASHTSSVKVENI  
NQDQLNLSRQLMGNSGANDQFWSVAHGPIFRVLALLPQAVLY

>EucaDof4

MIQELLGSPALIGERKISMNHGAVILEGMPSPPCFSPSPSPSSSTNSSNPENLRCP  
RCDSANTKFCYNNYNLTQPRHFCKTCRRYWTKGGALRNVPVGGGCRKNKSSAVSSSASK  
SSASKFKNLSSEIGRSSFVSAFDPEISQSQVLWASPQSSHILALLRANQNPNSLGHQA  
SPGTSTTKMVLSSVVKLKTGYNRYTKGSELPQVIIMVTVSTRRQF

>EucaDof5

MDSDWLQESESSGGSM LDSSTNSPSAADILAACSTRPQASAVAVAAAALMDGRRRLRPPP  
HHQPQKPCRCDSMNTKFCYNNYSLSQPRYFCKTCRRYWTHGGTLRNVVPGGGCRKAKRP  
SSNSSSTSSGKKPSNIVTANTS DLMALAHSHQNYQHSPLGFSHFGMMGYSYTPHEGN  
VGFLESKYGGLLSQSPRIDFLDSKFDLMGVNNDNLVMVNHGSNGDHHHHHHNHMGLNHG  
VGLNNNNNGGFNGISTGGNGGGGLMDISTCQRLMLSNDHHYHNEQEDHQRVATIMDV  
KPNPKLLSLDQQDQCYNNGGGGGAGKSDGGGYNGGGYINGLSSWNGLMNGYGTSTKT  
NSLV

>EucaDof6

MSEPKDQAFKFLFGKTIQVPEISVTTATTTDDDDDDSDQDRPSCANSSLDENITDDYN  
NNDKRDHGEEDTETDDKDSVGKTTIENQEDGASPVAAKESNLDATSGTSENPRTPSVDK  
ESTTPRTSKSEEEQSDTSNSQEKVLKPKDILPCPCNSMDTKFCYNNYNVQPRHFCK  
NCQRYWTAGGTRNVPV GAGRRKNKNSASHYRHLAVPEALQSVRTDFPNGVHHPPLKPNG  
TVLTFGSDAPLCEMASVLHLADKAMHNSRKNGFHKPEGVIMSPCGSVRSVETMNGSS  
ATAANSNLVESKNGTKEIVMPNCQGFPPQLPCFPGAPWPYAWNSAPWSPPISSLVHLVLL  
CRFTQGLIGVAVCQVRGTFLGLPSQLLRTTAPRAPHILRPWGNTRGMRKSGNQTTPLK

RSQRRLLLRDPCGFQRH

>EucaDof7

MERAWNPSVEMSPNCPRCGSSNTKFCYNNYSLTQPRYFCKGCRRYWTKGGSLRNVPVGG  
GCRKNRRGSLRLSSTDVAQSKSLGYGGVTADSTRHVHGYKGINPSIESCSSSTRSNGSHI  
DLALVYANFLNPQQPDKTASSGFEMQELGSDHFDPSLDFSGISNVNLDSTTVQLHDIGD  
HNFAGCIPAISSMGGDPPSDNHQLMHYCGVDSSSKHKQEANDHKIPQCTSITHDPINYG  
LPPLPGEEIGSQEILWPSSHMMESHTLQVTQQQAVLGPDQTQDPNTLLFGNWSFPDLSS  
DDAFSRS

>EucaDof8

MLEKRARPQEQNLNCPRCNSNNTKFCYNNYSLTQPRYFCKTCRRYWTGGSLRNVPVGGG  
SRKNKKVSTSSSSPSSASSRLPDLNPSSVISHFSAQNPKAGLLYEGQDLNLGFPQTQEY  
RATIAQYVQVPKVDPETSSRGPDNNSSAPSSALELLRNSSSGMGRGLSLPFLPANIQAD  
SNVLYSSAGFNLQELKPSLGFMSDRYASAHGCVVHQENGGKLLFPFGMMKQLSSTSGSD  
QVDHNGKGGSSNAYWNGMLGGGSW

>EucaDof9

MEETSPNACNNRAMLEKRARPQEQNLNCPRCNSNNTKFCYNNYSLTQPRYFCKTCRRYWT  
DGGSLRNVPVGGGSRKNKKVSTSSSSPSSASSRLPDLNPSSVISHFSAQNPKAGLLYEG  
QDLNLGFPQTQEYRATIAQYVQVPKVKTENRNNQHNSSSPYTSSPISAMELLRSGFAS  
RGLNTFIPTMPDSNTLYSSGGFPLQELRPTLSPADGLGSRYGQIENSGRLLFPFGELK  
QLSSTTSEVDQNKGGGSSGYWNGMFGGGSW

>EucaDof10

MEGSSRPALRSRPQKDQALNCPRCTSTHTKFCYNNYSLSQPRYFCKTCRRYWTEGGSLR  
NVPVGGGSRKNKRPSSSSPSTSSSSSKNKIIPNLTPPPPTILHSDPQNPNNKPQEGQGFG  
LPSYLPQAYNNSISEFADLSYNGDSKPHLQNPPTSSFSSHHLAMELLKSTGIASRGL  
GSFMPMSVSDSNISYSSGFPLQEFKPTLNFSLDGFQSGYGLQGVQESGARLLFPLEDLK  
QVSNTEFEQSRGVQGDSAGYWNGMLGGGSMVNNLEKKAIAIKKKKKKKVVKVGFSS  
LFFSFTCFADWLWVSR

>EucaDof11

MLSNCDEHEKMALLISSTTNEWPLQDDIQIDDIHHQKGLNNNNNNVRSSTSSPANRVME  
KPGQEQLQLQQPLKCPKCDSSNTKFCYNNYSLSQPRHFCKACKRYWTRGGTLRNVPVG  
GGCRKNKRVKRPASASATDTAPSSASATSDHHPQPHEFGPVLWVYGLSSNPFMDTA  
SGFDLHQLNHHAQLGLGFSSSGTNMSSDNYNRNFNPASSSSLPNSYSSMFSASPSTTV  
PTMASVLAASKFINGGAFKDIQARGPHQQNQYFQSLLPFEDLQIAAGNNGEVGGAADVKV  
EDHNGQNRQLQWNNNGSSNVNGNMNCQNQISEQMGLSDPTYWNSSTLGSNWHDPANLGS  
SVSSLI

>EucaDof12

MQEMGAHSLESMLISPKPQQDQRKPRPQPEALHCPRCESTNTKFCYNNYSLQPRYFC  
KSCRRYWTKGGTLRNVPVGGGCRKNRRSSSSSSRRADQPLMSNTSPLTNFPPLSYDT  
NDLSLAFARLQKQGGQLGFDDQHDFSILGNPVSSHCDILGNPSSSSTSATTTPAFLDA  
LRSGFLDTHNNFQNL

>EucaDof13

MDPSSQMSRQSSMENMLVCSKEQESKPKPQPEALHCPRCESTNTKFCYNNYSLQPR  
YFCKSCRRYWTKGGTLRNVPVGGGCRKNRRSSSSSSRRIQDQAFAPNHNPFNLPHFDQ  
SNDFALRLARLQKQSCGGMGYDENLSMLGNSGSSMNQSMSINHGFMDAIRSGLFLGNG



MHYNNSVQNMYYGGGYGNGDNGEVNSGNNCCGVSEEMMFNYDQEMSHCNAAINMKQESES  
RVFGGFPWQINGGIGNMGEVEQARASWNNGFTTPSWQGLLHSPLM

>EucaDof14

MAEAKDPAIKLFGKTIPLPEVATAAAGNDSPSGATVGGGGEDWVDQNRATNSSPEEDCVR  
AGEEGREVDKDTSGGKVTDRQEDGARSSSTSEEFDPDANSVGNENPKTPSADKETATLK  
CSKNEEQSETSISQEKTLKKPKDKILPCPRCNSMDTKFCYNNYNNVQPRHFCKNCQRYW  
TAGGTMRNVPVGAGRRKKNKNSHSHYRHITVSEALQSAGMFLWVLAARTRTRHLITVMDF  
PNGVHHPPLKPNGTVLTFGSDAPLCEMASVHLADKAMHNSTVLSGGSLALCMEFCSME  
PANIPFCPPGFVAVPFYAGPAYWSCSVPGAWNIPWVAQTPSPNHSATSSAPNSPTLGKHSR  
DEKIGKPNDSTEEPKEANPERSLWIPKTLRIDDPEAAKSSIWATLGIKNEASNSAGRG  
VLLKAFQSKSEEKNHVVEASSVLQANPAALSRSRNFQE

>EucaDof15

MSEPKDQAFKLFKTIQVPEISVTATTDDDDDDSDQDRPSCANSSLEDETITDDYN  
NNDKRDHGEEDTETDDKDSVGKTTIENQEDGASPVAAKESNLDATSGTSENPRTPSVDK  
ESTTPRTSKSEEQSDTSNSQEKVLKKPKDKILPCPRCNSMDTKFCYNNYNNVQPRHFCK  
NCQRYWTAGGTMRNVPVGAGRRKKNKNSASHYRHILAVPEALQSVRTDFPNGVHHPPLKPNG  
TVLTFGSDAPLCEMASVHLADKAMHNSRKNKNGFHKPEGVIMSPCGSVRSGVETMNGSS  
ATAANSNLVESKNGTKEIVMPNCQGFPPQLPCFPGAPWPYAWNSAPWSPPISSLVHLVLL  
CRFTQGQLIGVAVCQVRGTFGLPSQLLRTTAPRAPHILRPWGNTRGMRKSGNQTTPLK  
RSQRRLILRDPGCFQRH

>EucaDof16

MQDLTSAAYYHQSMMMTTQLKPCRCSTNTKFCYNNYNNLSQPRHFCKSCRRYWTGGA  
LRNIPVGGGTRKTAKRASSKSNLKNQTVAEKPDHHGSGSEEKEERVSGQEMNPTRMLYG  
LPVGDPNGASFSSLLASNMQMGLVYESGSRWLPMDLGLGSVRRSDDTWTDLAMNRMEK  
N

>EucaDof17

MQDIHDFSMNGVGGGGGGGRFFGGGGGGGDRRMRAHQNNILNHHQSLKPCPRCNSLN  
TKFCYNNYNNLSQPRHFCKNCRRYWTGGVLRNVPVGGGCRKAKRSKTSIGIPGTHASAH  
ASGRKEGREGGRKRGRRRRTCKISTRSQNVLDLPTSTSQGNRPGGGGGGGGGFGALDW  
QPGADQGLFDLPNAVHDHAYWSQSQWSTDGDQDPTGLYLP

>EucaDof18

MPAELSSSEAAARRAQSTAATLTKPGGAPPQEHLPCPRCDSTNTKFCYNNYNNFSQPR  
HFCKSCRRYWTHGGTLRDPVGGGTRKNAKRSRTSATSPAFMGPMGTNIDGLPLPATP  
VLLPLTANQGLSVHFGGGDGKNGGGGAGVGSNGGLGGSFTSLNLTQGPAGFLALNGFGLG  
IGPGIEDVNFALGRNAWQFEGGDEGLVGGGGGGSGGGGDSCFSWPGLAISTAWSGLK

## Conclusões e Perspectivas

### Obtenção das Sequências Dof de *Eucalyptus grandis*

O trabalho realizado resultou na descrição de 18 sequências gênicas potencialmente codificadoras de proteínas da família Dof de *E. grandis*, que irão enriquecer bancos de dados de proteínas e fatores de transcrição de plantas, podendo ser utilizadas em futuros trabalhos de análise (filo)genéticas.

O alinhamento demonstrou a presença do domínio Dof conservado, o que é uma forte evidência da funcionalidade destas seqüências. A construção de árvores filogenéticas permitiu classificar as proteínas em grupos de parálogos e ortólogos com proteínas de outra duas espécies: *A. thaliana* e *P. trichocarpa*. Os baixos valores de *bootstrap* encontrados nas árvores foram provavelmente devidos à baixa identidade das proteínas da família Dof fora do domínio principal. A localização das proteínas nas árvores filogenéticas (organização das seqüências com alta identidade protéica em ramos próximos) foi reforçada pela análise posterior com a ferramenta MEME, que demonstrou a presença de motivos conservados fora do domínio Dof dentro das mesmas. Entretanto, é importante observar que nem todas as proteínas classificadas como ortólogas na segunda árvore foram confirmadas com o alinhamento do MEME. Essa disparidade pode ser atribuída aos baixos valores de *bootstrap* encontrados na árvore.

A presença de motivos conservados e classificação de grupos de ortólogos irão auxiliar em futuros estudos de função dessas proteínas.

### Realização do Trabalho

A importância das análises aqui realizadas pode ser considerada alta, visto que se trata de um trabalho inédito de anotação gênica. Entre as maiores dificuldades encontradas na realização desta monografia esteve a obtenção das sequências completas dos genes, e a utilização do *software* de descoberta de motivos conservados MEME, por utilizar uma plataforma não familiar ao aluno (Linux).

## **Análise da Expressão das Proteínas Dof**

Como próximo passo no trabalho, será realizada a análise da expressão das proteínas Dof anotadas em diferentes tecidos de *E. grandis* sob diferentes tratamentos. Para tanto, será extraído o RNA total de flor, folha, xilema, e plântulas tratadas com frio, água, seca, ácido abscísico (ABA), ácido naftalenoacético (NAA), quinetina e NaCl.

O cDNA total dessas amostras será obtido por meio de experimentos de PCR precedida de transcrição reversa com a enzima transcriptase reversa (RT-PCR). A análise de expressão se dará por meio de experimentos de PCR quantitativo (qPCR), utilizando *primers* desenhados a partir das 18 sequências de proteínas Dof anotadas. A análise da expressão das proteínas sob diferentes tecidos e tratamentos irá fornecer importantes informações sobre suas funções e locais de ação.