

# “DATAMINING” DOS GENES DA CELULOSE SINTASE RELACIONADOS COM ESTs DE *Eucalyptus* spp. (NOTA CIENTÍFICA)<sup>1</sup>

Léo ZIMBACK<sup>2</sup>  
Edson Seizo MORI<sup>3</sup>  
Mario Luiz Teixeira de MORAES<sup>4</sup>  
Edson Luiz FURTADO<sup>5</sup>  
Carlos Frederico WILCKEN<sup>6</sup>  
Iraê Amaral GUERRINI<sup>7</sup>  
Edivaldo Domingues VELINI<sup>8</sup>  
Celso Luis MARINO<sup>9</sup>  
Daniel Dias ROSA<sup>10</sup>  
Hideyo AOKI<sup>11</sup>

## RESUMO

Trata-se de um estudo sobre “datamining” envolvendo genes ligados ao crescimento de controle não hormonal, utilizando o banco de dados de ESTs de eucalipto, efetuado através do Projeto Genoma do Eucalipto (FORESTs) comparados ao nível de aminoácidos. Foram identificados os clusters de ESTs EGBGFB1211D01.g, EGEZRT6201E10.g, EGCCFB1220G07.g, EGRFCL1206E01.g, EGEQST2006A06.g, EGRFCL1206E01.g, EGEQRT3001H05.b e EGBFRT3106G11.g, similares às proteínas de celulose sintase e suas subunidades controlando o crescimento em *Arabidopsis thaliana*, *Gossypium hirsutum*, *Populus tremuloides*, *Zea mays* e *Nicotiana glauca*, registradas no National Center of Biotechnologies Information - NCBI, informação valiosa para futuros programas de melhoramento genético do gênero *Eucalyptus*.

Palavras-chave: *Eucalyptus* spp.; genes de crescimento; datamining; EST.

## 1 INTRODUÇÃO

Os caracteres de crescimento em plantas são influenciados pela atuação de centenas de enzimas, envolvidas em vários passos metabólicos, tais como fotossíntese, respiração, síntese de carboidratos,

## ABSTRACT

This is a study about data mining of expressed sequence tags (ESTs) involved with cellulose synthase growth effect genes resulted from the Eucalyptus ESTs Genome Project (FORESTs) compared at aminoacids level. Using a sequencing of derived from cDNAs libraries induced and not induced by bacteria, were identified EST clusters EGBGFB1211D01.g, EGEZRT6201E10.g, EGCCFB1220G07.g, EGRFCL1206E01.g, EGEQST2006A06.g, EGRFCL1206E01.g, EGEQRT3001H05.b, and EGBFRT3106G11.g, similar to cellulose synthase proteins controlling growth effect in *Arabidopsis thaliana*, *Gossypium hirsutum*, *Populus tremuloides*, *Zea mays*, and *Nicotiana glauca*, registered on National Center of Biotechnologies Information - NCBI. These mining results are important to improve *Eucalyptus* breeding programs.

Key words: *Eucalyptus* spp.; growth genes; datamining; TAG.

aminoácidos e produtos do metabolismo secundário. Até mesmo a resistência ao estresse, a pragas e doenças, florescimento, produção de sementes e morfologia de folhas, interferem no crescimento, por afetarem indiretamente alguns passos metabólicos.

(1) Aceito para publicação em fevereiro de 2008.

(2) Instituto Florestal, Floresta Estadual de Botucatu, 18610-120, Jardim Mirante, Botucatu, SP, Brasil. E-mail: lzimback@terra.com.br

(3) Faculdade de Ciências Agronômicas da UNESP, Caixa Postal 237, 18603-970, Botucatu, SP, Brasil. E-mail: esmori@fca.unesp.br

(4) Escola de Engenharia de Ilha Solteira, UNESP, Caixa Postal 322, 15.385-000, SP, Brasil. E-mail: teixeira@agr.feis.unesp.br

(5) Faculdade de Ciências Agronômicas da UNESP, Caixa Postal 237, 18603-970, Botucatu, SP, Brasil. E-mail: elfurtado@fca.unesp.br

(6) Faculdade de Ciências Agronômicas da UNESP, Caixa Postal 237, 18603-970, Botucatu, SP, Brasil. E-mail: secddf@fca.unesp.br

(7) Faculdade de Ciências Agronômicas da UNESP, Caixa Postal 237, 18603-970, Botucatu, SP, Brasil. E-mail: iguerrini@fca.unesp.br

(8) Faculdade de Ciências Agronômicas da UNESP, Caixa Postal 237, 18603-970, Botucatu, SP, Brasil. E-mail: secdamv@fca.unesp.br

(9) Instituto de Biociências, UNESP, Caixa Postal 456, 18603-970, Botucatu, SP, Brasil. E-mail: clmarino@ibb.unesp.br

(10) Discente da Faculdade de Ciências Agronômicas da UNESP, Caixa Postal 237, 18603-970, Botucatu, SP, Brasil. E-mail: DNascimento@sial.com

(11) Instituto Florestal, Floresta Estadual de Avaré, Caixa Postal 78, 18701-180, Avaré, SP, Brasil. E-mail: hiaoki@iflorestal.sp.gov.br

A redução da síntese da parede primária da fibra é associada a um decréscimo na elongação celular que é um fator importante no crescimento, como é observado em mutantes *cev1* em *Arabidopsis* com aumento da síntese de jasmonato e etileno (Ellis *et al.*, 2002). Os genes de celulose sintase *AtCesA1*, *AtCesA3* e *AtCesA6* de *Arabidopsis* também estão envolvidos no desenvolvimento da parede primária da célula, com diferenças de níveis na expressão de cada gene no crescimento, enquanto *AtCesA4*, *AtCesA7* e *AtCesA8* estão envolvidos na síntese da parede secundária da célula (Joshi, 2003), possuindo alta similaridade com os genes de celulose sintase da espécie arbórea *Populus tremuloides* (Kalluri & Joshi, 2003), fato útil em futura manipulação genética. O gene *PtCesA4* de *Populus*, ligado ao crescimento do xilema, possui também alta similaridade com o gene *GhCesA* do algodão (Wu *et al.*, 2000).

Um alelo mutante dominante do gene *AtCesA7* provoca fragilidade da parede celular (Zhong *et al.*, 2003) e em *Arabidopsis* observou-se mutante do gene *CesA1* (Burn *et al.*, 2002) alterando o comprimento da fibra mostrando que existe variabilidade para a expressão destes genes e que pode ser explorada no melhoramento. A identificação de novos cDNAs de genomas de árvores é essencial para a compreensão do mecanismo da síntese da celulose destas. Tal conhecimento pode ser empregado para uma seleção mais eficiente em eucalipto, objetivando-se ganhos na taxa de crescimento para fins comerciais.

Este estudo teve por objetivo identificar no Banco de ESTs de *Eucalyptus* (FORESTs da FAPESP), aqueles que apresentam alta homologia com os genes da celulose sintase que estão depositados no NCBI (2004).

## 2 MATERIAL E MÉTODOS

O Eucalypt Genome Project (FORESTs) (2004) foi desenvolvido pelas companhias florestais Ripasa S/A, Votorantim Celulose e Papel, Suzano Bahia Sul e Duratex S/A associadas

com a Fundação de Amparo à Pesquisa do Estado de São Paulo - FAPESP.

O seqüenciamento realizado por 20 laboratórios da Rede Genoma do Estado de São Paulo permitiu a obtenção de 123.889 leituras de "expressed sequence tags" (ESTs) agrupados em clusters de seqüências de alta homologia, *E. grandis*, cujos dados estão no AEG network (Eucalypt Genome Project, 2004). As bibliotecas de cDNAs foram obtidas de madeira (WD), casca (BK), gema floral (FB), folhas (LV), plântulas (SL), ramos (ST), raízes (RT) e calos (CL), em diferentes condições de crescimento e ambientes.

A busca das proteínas que teriam homologia com os ESTs foi realizada com o programa BLAST de bioinformática (Altschud *et al.*, 1997), que gera valores de probabilidade das seqüências serem homólogas (e-value). Cada seqüência de nucleotídeos do EST foi comparada com as seqüências de proteínas depositadas no NCBI ou GeneBank. Árvore de similaridades foi desenhada através do método do vizinho mais próximo, utilizando-se o programa Clustal-X gerando 1000 bootstraps para análise das probabilidades de consistência dos nós, e visualizado pelo programa Tree View.

## 3 RESULTADOS E DISCUSSÃO

Na busca de similaridades dos clusters de ESTs de eucalipto (FORESTs) com o banco de proteínas (NCBI) para genes de crescimento, localizou-se o gene *Ath-A* (acesso NM\_120095.1 no NCBI, 2004) de celulose sintase e suas subunidades catalíticas apresentando efeitos no crescimento de *Arabidopsis thaliana*, o qual está associado à elongação celular. A TABELA 1 indica os resultados da busca considerando níveis de e-value abaixo de  $e^{-80}$ , onde a proteína NM\_120095.1 foi estreitamente relacionada com os clusters de eucalipto com e-value igual a 0.

TABELA 1 – Os melhores e-values encontrados na busca de proteínas em banco de dados associadas a clusters de *Eucalyptus* do Projeto Forests.

Organismo	Gene	Acesso	Clusters de <i>Eucalyptus</i>	e-value
<i>Arabidopsis thaliana</i>	Ath-A	gil30692282	EGEZRT6201E10.g	0,0
<i>Arabidopsis thaliana</i>	Ath-A	gil30692282	EGRFCL1206E01.g	0,0
<i>Arabidopsis thaliana</i>	Ath-A	gil30692282	EGCCFB1220G07.g	0,0
<i>Arabidopsis thaliana</i>	Ath-A	gil30692282	EGEQST2006A06.g	0,0
<i>Arabidopsis thaliana</i>	Ath-A	gil30692282	EGEQRT3001H05.b	0,0
<i>Arabidopsis thaliana</i>	Ath-A	gil30692282	EGEQFB1003E06.g	0,0
<i>Arabidopsis thaliana</i>	Ath-A	gil30692282	EGBGFB1211D01.g	e -172
<i>Arabidopsis thaliana</i>	Ath-A	gil30692282	EGJFFB1118A11.g	e -117
<i>Arabidopsis thaliana</i>	Ath-A	gil30692282	EGBFRT3106G11.g	e -101

Embora a maior similaridade tenha sido com o gene de celulose sintase de *Arabidopsis* (TABELA 1), a árvore de similaridades obteve associações com genes de celulose sintase de outras espécies, como se pode observar na FIGURA 1. Pear *et al.* (1996) obtiveram similaridade da seqüência de nucleotídeos de celulose sintase subunidade A de *Gossypium hirsutum* (acesso gil7484715) com a da seqüência de nucleotídeos de *Oriza sativa*, que também é associado ao cluster EGEZRT6201E10.g de eucalipto deste trabalho (FIGURA 1). Existem boas associações com outros acessos de *A. thaliana* como: EGCCFB1220G07.g para gil15224726, EGRFCL1206E01.g para gil15237958, EGEZRT6201E10.g para gil30694433 e EGEQST2006A06.g para gil15242540, sendo estes clusters obtidos de gemas florais, calos, raiz e ramos, respectivamente, indicando que este gene se expressa em diferentes tecidos e órgãos da espécie estudada.

A seqüência da espécie arbórea *Populus tremuloides* (Wu *et al.*, 2000) acesso gil23534479, que está na mesma chave do gene AtCesA7 (irx3) de *Arabidopsis*, acesso gil15237958 (Samuga & Joshi, 2002) de celulose sintase subunidade A, foi estreitamente relacionada ao cluster de calos EGRFCL1206E01.g na FIGURA 1. Em outra chave pode-se observar seqüência de *Populus* (gil4115905) e de *Arabidopsis* (gil18415170)

associados a um cluster obtido de raiz (EGEQRT3001H05.b). Da mesma forma, ocorreu alta similaridade de seqüências de celulose sintase obtidas de *Zea mays* por Holland *et al.* (2000) com clusters de eucalipto (gil96222878 e gil9622882 com EGCCFB1220G07.g e gil9622888 com EGEQFB1003E06.g). Outra espécie obtida na árvore de similaridades foi *Nicotiana glauca* (gil12925881 associado com o cluster EGBFRT3106G11.g) cuja seqüência foi estudada por Doblin *et al.* (2001). Assim, seqüências de celulose sintase subunidade A de cinco espécies vegetais foram associadas aos clusters do projeto FORESTS, não apresentando similaridades com clusters de folha, plântula, casca e madeira, mas a clusters de raiz, ramos, calos e gemas florais.

Os resultados mostram que esses ESTs possuem alta probabilidade de corresponder ao mRNA da enzima celulose sintase ou outra proteína com seqüência muito similar, mas como foram identificados em órgãos com atividade de alongação celular, como caule e raiz, é muito provável a identificação positiva em ensaios de validação. Como a variabilidade genética do gene da celulose sintase afeta a expressão do crescimento em algumas espécies, esses ESTs poderão ser utilizados em seleção genética assistida para o caráter em programas de melhoramento.

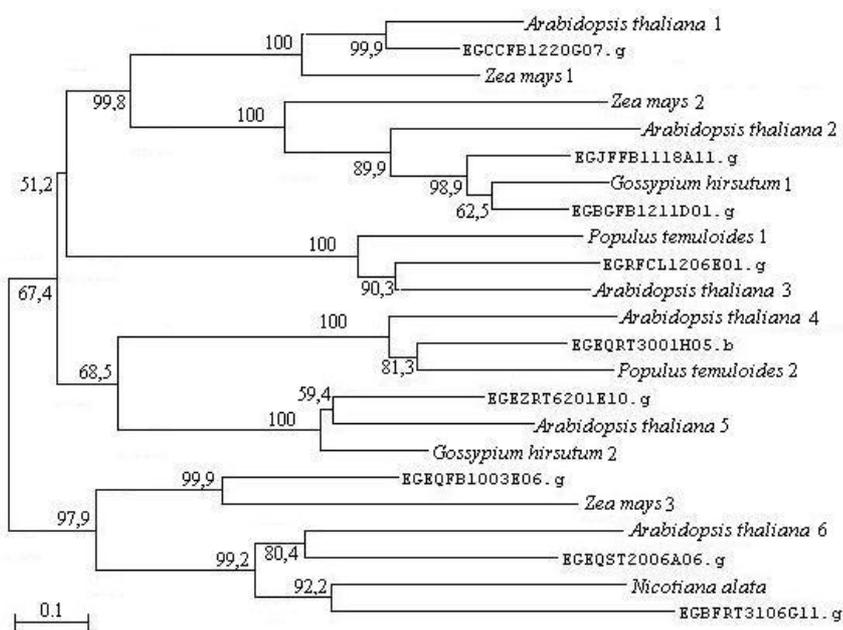


FIGURA 1 – Dendrograma de similaridade de seqüências de aminoácidos da celulose sintase e subunidades catalíticas. As seguintes seqüências foram obtidas do EMBL GeneBank: *Arabidopsis thaliana* (1 gil15224746, 2 gil9622882, 3 gil15237958, 4 gil18415170, 5 gil30694433, 6 gil15242540), *Zea mays* (1 gil9622878, 2 gil9622882, 3 gil9622888), *Gossypium hirsutum* (1 gil6446577, 2 gil7484715), *Populus tremuloides* (1 gil23534479, 2 gil4115905), *Nicotiana glauca* (gil13925881), as quais foram comparadas com os clusters de *Eucalyptus* (EG\*). A escala é corrigida para múltiplas substituições. O número indica a consistência probabilística do nó, sendo que a barra indica produto de 0,1 substituições do total.

#### 4 CONCLUSÕES

A busca de alelos mutantes do gene da celulose sintase e subunidades catalíticas com ação no crescimento de *Arabidopsis thaliana*, *Gossypium hirsutum*, *Populus tremuloides*, *Zea mays* e *Nicotiana glauca*, no banco de proteínas do NCBI, mostraram altas similaridades genéticas com os clusters de ESTs de eucalipto EGBGFB1211D01.g, EGEZRT6201E10.g, EGCCFB1220G07.g, EGRFCL1206E01.g, EGEQST2006A06.g, EGRFCL1206E01.g, EGEQRT3001H05.b e EGBFRT3106G11.g do projeto FORESTs.

#### 5 AGRADECIMENTOS

Os autores agradecem às companhias florestais privadas Ripasa S/A, Votorantim Celulose e Papel, Suzano Bahia Sul e Duratex S/A e à FAPESP pelo apoio financeiro e acesso ao banco de dados do FORESTs.

#### REFERÊNCIAS BIBLIOGRÁFICAS

- ALTSCHUD, S. F. *et al.* Grapped BLAST and PSI-BLAST: a new generation of protein database search programs. **Nucleic Acids Research**, Oxford, v. 25, p. 3389-3402, 1997.
- BURN, J. E. *et al.* Functional analysis of the cellulose synthase genes Cesa1, Cesa2, and Cesa3 in *Arabidopsis*. **Plant Physiology**, Rockville, v. 129, p. 797-807, 2002.
- DOBLIN, M. S. *et al.* Pollen tubes of *Nicotiana glauca* express two genes from different beta-glucan synthase families. **Plant Physiology**, Rockville, v. 125, p. 2040-2052, 2001.
- ELLIS, C. *et al.* The *Arabidopsis* mutant *cev1* links cell wall signaling to jasmonate and ethylene responses. **The Plant Cell**, Rockville, v. 14, p. 1557-1566, 2002.

ZIMBACK, L. *et al.* "Datamining" dos genes da celulose sintase relacionados com ESTs de *Eucalyptus* spp. (Nota Científica).

EUCALYPT GENOME PROJECT (FORESTs).

Disponível em: <<https://forests.esalq.usp.br>>. Acesso em: 25 abr. 2004 a 12 jun. 2004.

HOLLAND, N. *et al.* A comparative analysis of the plant cellulose synthase (CesA) gene family. **Plant Physiology**, Rockville, v. 123, p. 1313-1324, 2000.

JOSHI, C.P. Xylem-specific and tension stress-responsive expression of cellulose synthase genes from aspen trees. **Applied Biochemistry and Biotechnology**, Clifton, v. 105, p. 17-23, 2003.

KALLURI, U. C.; JOSHI, C. P. Isolation and characterization of a new, full-length cellulose synthase cDNA, PtrCesA5 from developing xylem of aspen trees. **Journal of Experimental Botany**, 2003, Oxford, v. 54, n. 390, p. 2187-2188.

NATIONAL CENTER OF BIOTECHNOLOGY INFORMATION - NCBI. Disponível em: <<http://www.ncbi.nlm.nih.gov>>. Acesso em: 25 abr. 2004 a 12 jun. 2004.

PEAR, J. R. *et al.* Higher plants contain homologs of the bacterial celA genes encoding the catalytic subunit of cellulose synthase. **Proc Natl Acad Sci**, Washington, D.C., v. 93, p. 12637-12642, 1996.

SAMUGA, A.; JOSHI, C. P. A new cellulose synthase gene (PtrCesA2) from aspen xylem is orthologous to Arabidopsis AtCesA7 (irx3) gene associated with secondary cell wall synthesis. **Gene**, Amsterdam, v. 296, n. 1-2, p. 37-44, 2002.

WU, L.; JOSHI, C. P.; CHIANG, V. L. A xylem-specific cellulose synthase gene from aspen (*Populus tremuloides*) is responsive to mechanical stress. **The Plant Journal**, Oxford, v. 22, p. 495-502, 2000.

ZHONG, R. *et al.* Expression of a mutant form of cellulose synthase AtCesA7 causes dominant negative effect on cellulose biosynthesis. **Plant Physiology**, Rockville, v. 132, p. 786-795, 2003.