

**ANÁLISE IN SILICO DO GENE *FRIGIDA (FRI)* EM CAFÉ ARÁBICA**

HORLLYS GOMES BARRETO<sup>1</sup>, ANDERSON TADEU SILVA<sup>2</sup>, SOLANGE APARECIDA SÁGIO<sup>3</sup>, ANTONIO CHALFUN JUNIOR<sup>4</sup>, LUCIANO VILELA PAIVA<sup>5</sup>.

**RESUMO**

Sincronizar o florescimento e amadurecimento dos frutos pode contribuir para reduções no custo de produção. A compreensão e identificação do mecanismo de florescimento, assim como os genes envolvidos nos processos de regulação floral, são de fundamental importância para elucidar como esses mecanismos ocorrem em café. Hoje com o seqüenciamento do transcriptoma do cafeeiro e a criação do banco CAFEST de ESTs (*Expressed Sequence Tags*) tornou-se possível a realização de buscas por seqüências formadoras dos prováveis genes relacionados às características de interesse. Dessa forma, estudos comparativos de genômica funcional surgem como ferramentas poderosas para a identificação de genes envolvidos na regulação de vias metabólicas. Diante disso, o presente trabalho teve como objetivo identificar o gene *FRI* em café arábica, gene esse já descrito em *Arabidopsis*, que está envolvido no complexo de regulação gênica do florescimento, pois, esse trabalho abre caminho para a identificação de genes de interesse para estudos aplicados que possam melhorar a cadeia produtiva do café, o que pode permitir um controle do tempo de florescimento e aumento da produtividade. Através das ferramentas da bioinformática foi possível a identificação de dois EST-*contigs* e um *singlet* que são fortes candidatos para o gene *FRI* em café arábica.

**Palavras-chaves:** Bioinformática, Florescimento, Expressão gênica, Regulação floral.

**INTRODUÇÃO**

O cafeeiro (*Coffea sp.*) pertence a família *Rubiacea* é responsável pela produção da segunda bebida mais consumida pela população brasileira, o café, perdendo apenas da água (TNS InterScience, 2008). O Brasil ocupa o primeiro lugar na produção e exportação de café, com 47,274 mil sacas de 60 kg/ano e 34,28% respectivamente, com área plantada de 2.101.097 ha e produtividade média de 24,54 sacas/ha (CONAB, 2010). Em relação à produção, mesmo o cafeeiro sendo uma cultura bastante estudada, ainda existe problemas com a produtividade, reflexo principalmente do florescimento seqüencial, que causa uma maturação desuniforme dos frutos, dificultando a colheita e causando perdas de produção. Sincronizar o florescimento e amadurecimento dos frutos pode contribuir para reduções no custo de produção. A compreensão e identificação do mecanismo de florescimento, assim como os genes envolvidos nos processos de regulação floral, são de fundamental importância para elucidar como esses mecanismos ocorrem em café. Vários genes já foram identificados como peças chaves na regulação molecular do florescimento. Em função do seu pequeno genoma e do ciclo curto de vida, a planta *Arabidopsis thaliana* tem sido utilizada como modelo para entendimento das diferentes rotas metabólicas, incluindo aspectos encontrados no processo de floração (LEVY et al., 2002; SIMPSON & DEAN, 2002). No momento, toda complexidade da rede de florescimento está baseada em informações importantes já obtidas em *Arabidopsis* (IZAMA et al., 2003). Um dos principais *locus* responsável pela regulação do florescimento em *Arabidopsis* é o FRIGIDA (*FRI*) que regula a expressão do maior repressor floral o *FLOWERING LOCUS C (FLC)*, pois quando ativo conduz a ativação de genes de identidade de meristemas (SHINDO et al., 2005, CIANNAMEA, 2007). Hoje com o seqüenciamento do transcriptoma do cafeeiro (VIEIRA et al., 2006) e a criação do

---

<sup>1</sup> Mestrando em Biotecnologia Vegetal, LCBM/UFLA, horllys@posgrad.ufla.br

<sup>1</sup> Mestrando em Biotecnologia Vegetal, LCBM/UFLA, anderson.tadeu@yahoo.com.br

<sup>3</sup> Doutoranda em Fisiologia Vegetal, DBI/UFLA, salsagio@gmail.com

<sup>4</sup> Professor Adjunto, DBI/UFLA, chalfunjunior@ufla.br

<sup>5</sup> Professor Associado, LCBM/UFLA, luciano@ufla.br

banco CAFEST de ESTs (*Expressed Sequence Tags*) tornou-se possível a realização de buscas por sequências formadoras dos prováveis genes relacionados às características de interesse. Dessa forma, estudos comparativos de genômica funcional surgem como ferramentas poderosas para a identificação de genes envolvidos na regulação de vias metabólicas. Diante disso, o presente trabalho teve como objetivo identificar o gene *FRI* em café arábica, gene esse já descrito em *Arabidopsis*, que está envolvido no complexo de regulação gênica do florescimento.

## **MATERIAL E MÉTODOS**

### **Análise *In Silico***

As sequências dos gene *FRIGIDA* (*FRI*) foram identificadas tendo-se como fonte de dados o banco de ESTs (*Expressed Sequence Tags*) gerado pelo projeto Genoma Brasileiro Café (CAFEST; VIEIRA et al., 2006). Por meio da interface *Gene Project* (<http://www.lge.ibi.unicamp.br>), foi possível a procura por *reads*, montagem de *clusters* e análise dos prováveis genes.

### **Análise Filogenética**

O alinhamento dos EST-contigs foi feito pelo programa ClustalW (THOMPSON et al., 1994) com os parâmetros padrões (default), utilizando-se as sequências de nucleotídeos traduzidas em aminoácidos. As sequências foram visualmente inspecionadas e manualmente corrigidas, sendo removidos segmentos cuja homologia não pode ser acertada. A árvore final foi feita utilizando-se o programa MEGA 4.0 (TAMURA et al., 2007), com o modelo de comparação *Neighbor-joining* (SAITOU & NEI, 1987).

### **Identificação de *motifs* comuns de agrupamento**

Para descobrir *motifs* de agrupamento entre as sequências do gene *FRIGIDA* (*FRI*) selecionadas no CAFEST, foi utilizado o programa MEME (*Multiple Expectation Minimization for Motif Elicitation*, [http://meme.sdsc.edu/meme4\\_4\\_0/intro.html](http://meme.sdsc.edu/meme4_4_0/intro.html)) versão 3.5.4 (BAILEY & ELKAN, 1994). A sequência total de aminoácidos das proteínas foi comparada de acordo com a ordem de agrupamento observada na árvore filogenética, facilitando dessa forma a observação dos *motifs* de agrupamento em cada subclasse. Os parâmetros utilizados foram: número de repetições qualquer, máximo número de *motifs* 20 e amplitude ótima entre 6 e 200 (PARENICOVA et al., 2003).

### **Northern eletrônico**

Para a análise espacial de expressão, foram calculadas as frequências de *reads* formadores de cada EST-*contig* nas bibliotecas em que apareceram expressos. Em seguida, os dados foram normalizados, para dar uma idéia exata do grau de expressão dos prováveis genes em cada tratamento e local da planta. O procedimento para a normalização consistiu em multiplicar cada *read* pela razão entre o número total de *reads* de todas as bibliotecas e o número de *reads* da biblioteca em que ele aparecia expresso. As sequências foram reorganizadas de acordo com os locais de expressão e seus dados foram lançados em uma matriz relacionando genes e bibliotecas. Os EST-*contigs* e bibliotecas foram agrupados por *hierachical clustering* utilizando-se os programas Cluster e TreeView (EISEN et al., 1998).

## **RESULTADOS E DISCUSSÃO**

A clusterização (agrupamento) realizada com os 57 *reads* encontrados no banco de dados CAFEST formou 6 EST-*contigs* e 7 *singlets*. Após a análise das sequências protéicas, da parte funcional de cada uma delas e de seus domínios conservados, observou-se que dois EST-*contigs* (1 e 5) e um *singlet* apresentaram o domínio *FRIGIDA* (*motif* 1) muito conservado (Figura 1). Os EST-*contigs* e *singlets* incompletos (sem o domínio) foram excluídos da análise e omitidos nos resultados.

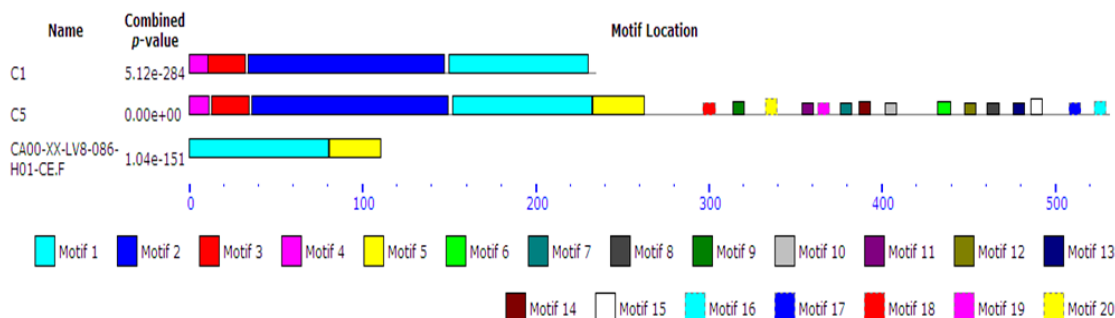


FIGURA 1 – Representação gráfica mostrando os motivos de agrupamento entre os prováveis EST-contigs para o gene *FRI* encontrados no CAFEST, utilizando-se o programa MEME (*Multiple Expectation Minimization for Motif Elicitation*, <http://meme.sdsc.edu/meme/meme.html>).

Com a submissão dos EST-contigs 1 e 5 e o singlet às análises filogenéticas em comparação com outras espécies, em que o gene *FRI* já foi identificado e caracterizado, observou-se que ambos agruparam no mesmo clado e apresentaram similaridade com espécies como *Medicago truncatula*, *Glycine max*, *Vitis vinifera* e *Arabidopsis thaliana* (Figura 2). Isso demonstra que foi possível selecionar fortes candidatos para o gene *FRI* em cafeeiro.

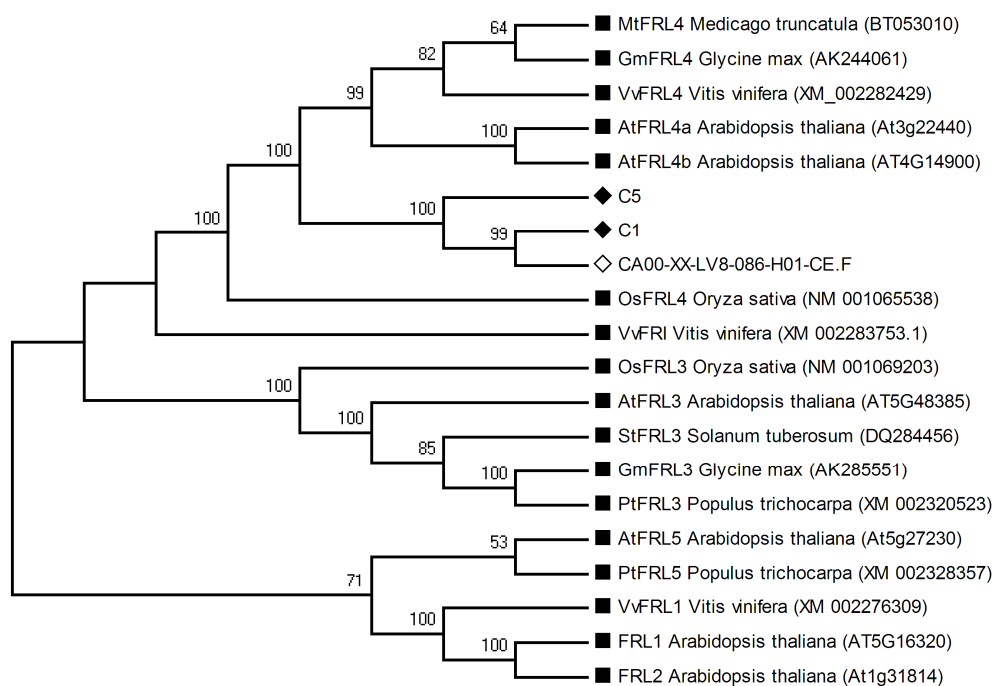


FIGURA 2 – Árvore filogenética relacionando as 8 prováveis sequências de *FRI* de *C. arabica* encontradas no CAFEST, e as proteínas de espécies pertencente à subfamília *FRI* publicadas e depositadas no NCBI. (♦) EST-contigs, (◇) singlets e (■) espécies já descritas. Foi utilizado o modelo de comparação *Neighbor-joining*, pelo método de distância *p* e supressão *pair-wise*. Valores de Bootstrap menores do que 50% foram omitidos.

Com análise do northern eletrônico, observou-se que o EST-contig 5 foi o que obteve maior padrão de expressão, pois o mesmo apresentou perfil de expressão na maior parte das bibliotecas em estudo, exceto na biblioteca de folhas maduras de ramos plagiotrópicos, porém a biblioteca de folhas jovens de ramos ortotrópicos foi uma das bibliotecas com maior perfil de expressão. (Figura 3).

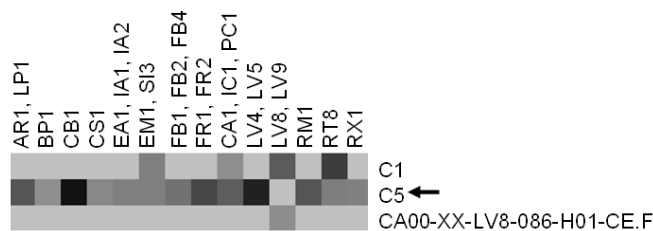


FIGURA 3 Northern Eletrônico representando os níveis de expressão dos EST-*contigs* nas diferentes bibliotecas por meio de uma escala cinza. Bibliotecas: AR1/LP1, plântulas e folhas tratadas com ácido araquidônico; BP1, células em suspensão tratadas com acibenzolar-S-metil; CB1, células em suspensão tratadas com acibenzolar-S-metil e brassinoesteroids. CS1, células em suspensão tratadas com NaCl; EA1/IA1/IA2, calos embriogênicos; EM1/SI3, sementes no início da germinação; FB1/FB2/FB4, botões florais em diferentes estágios do florescimento; FR1/FR2, botões florais e frutos de diferentes estágios; CA1/IC1/PC1, calos não embriogênicos; LV4/LV5, folhas jovens de ramos ortotrópicos; LV8/LV9, folhas maduras de ramos plagiotrópicos; PA1 calos embriogênicos primários; RM1, folhas infectadas com bicho mineiro do cafeeiro e ferrugem; RT8, células em suspensão com estresse ao alumínio; RX1, talos infectados com *Xylella spp.* As bibliotecas utilizadas foram descritas por Vieira et al., (2006).

## CONCLUSÃO

Foi possível através do uso da bioinformática identificar dois EST-*contigs* e um *singlet*, que são fortes candidatos para o gene *FRI* de cafeeiro.

## REFERÊNCIAL BIBLIOGRÁFICO

BAILEY, T.L.; ELKAN, C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. In: INTERNATIONAL CONFERENCE ON INTELLIGENT SYSTEMS FOR MOLECULAR BIOLOGY, 2., 1994, Menlo Park. **Proceedings...** Menlo Park: AAI, p. 28–36. 1994

CIANNAMEA, S. **Molecular characterization of the vernalization response in *lilium perenne*.** 2007. 147p. Tese (Doctor in Plant Science) – Wageningen University, Wageningen.

COMPANHIA NACIONAL DE ABASTECIMENTO. **Primeira estimativa:** janeiro 2010. Disponível em: < [www.abic.com.br/estat\\_pagricola.html](http://www.abic.com.br/estat_pagricola.html) >. Acesso em: 20 jan. 2010.

EISEN, M.B.; SPELLMAN, P.T.; BROWN, P.O.; BOTSTEIN, D. Cluster analysis and display of genome-wide expression patterns. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v.95, n.25, p.14863-14868, Nov. 1998.

IZAWA T, TAKAHASHI Y, YANO M Comparative biology comes into bloom: genomic and genetic comparison of flowering pathways in rice and Arabidopsis. **Curr Opin Plant Biol** v.6, p113–120. April 2003.

LEVY, Y.Y.; MESNAGE, S.; MYLNE, J.S.; GENDALL, A.R.; DEAN, C. Multiple roles of *Arabidopsis* VRN1 in vernalization and flowering time control. **Science**, Washington, v.297, n. 5579, p.243-246, July 2002

PARENICOVA, L., et al. Molecular and phylogenetic analyses of the complete MADS-box transcription factor family in Arabidopsis: New openings to the MADS world. **The Plant Cell**, Rockville, v. 15, n.7, p.1538-1551, July 2003.

**XIX CONGRESSO DE PÓS-GRADUAÇÃO DA UFLA**  
**27 de setembro a 01 de outubro de 2010**

---

SAITOU N, NEI M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. **Mol Biol Evol** v.4, p.406–425. Jul. 1987

SIMPSON, G. G., DEAN, C. *Arabidopsis*, the Rosetta stone of flowering time? **Science**, v.296, p.285-289, April 2002.

SHINDO, C. et al. Role of FRIGIDA and FLOWERING LOCUS C in determining variation in flowering time of *Arabidopsis*. **Plant Physiology**, Rockville, v.138, n.2, p.1163- 1173, June 2005.

TAMURA, K.; DUDLEY, J.; NEI, M.; KUMAR, S. H. MEGA4: molecular evolutionary genetics analysis: software: version 4.0. **Molecular Biology and Evolution**, Oxford, v. 24, n. 8, p.1596-1599, Aug. 2007.

TNS RESEARCH INTERNATIONAL. **Primeira estimativa** 2008: janeiro de 2008  
<<http://www.abic.com.br/estatisticas.html#intro>>. Acesso em 20 jan 2010

THOMPSON, J.D.; HIGGINS, D.G.; GIBSON, T.J.; CLUSTAL, W. Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. **Nucleic Acids Research**, Oxford, v.22, n.22, p.4673-4680, Nov. 1994.

VIEIRA L.G.E. et al. Brazilian coffee genome project: an EST-based genomic resource. **Brazilian Journal Plant Physiology**, Londrina, v.18, n.1, p.95-108, Jan./Mar. 2006