

# EXPRESSÃO DE FATORES DE TRANSCRIÇÃO TIPO WRKY NO BANCO GENOMA FUNCIONAL DE CAFÉ (*Coffea sp.*)

Maria Cristina R. CORDEIRO<sup>1</sup>, E-mail: [cristina@cpac.embrapa.br](mailto:cristina@cpac.embrapa.br); Marília S. SILVA<sup>1</sup>; Natália F. MARTINS<sup>2</sup>; Maria Fátima GROSSI DE SA<sup>2</sup>

<sup>1</sup>Embrapa Cerrados, BR 0 20, km 18, CEP 70310-970, <sup>2</sup>Embrapa Cenargen, Est. P.Rural, CEP 70849-070.

## Resumo:

A expressão de genes relacionados a fatores de transcrição tipo wrky no genoma funcional do café foi avaliada. Foram observadas 227 reads relacionadas com proteínas WRKY no *Coffea arabica*, 205 reads no *Coffea canephora* e 10 reads no *Coffea racemosa*. Dentre as seqüências que apresentaram homologia com proteínas WRKY encontram-se seqüências de proteínas que apresentam o domínio NBS-LRR em sua estrutura assim como proteínas relacionadas a genes de resistência de espécies diferentes. As seqüências obtidas estão relacionadas a diferentes situações fisiológicas em especial desenvolvimento, estresse biótico e abiótico.

Palavras-chave: Genoma funcional, café, WRKY

## EXPRESSION OF WRKY TRANSCRIPTION FACTORS WITHIN THE COFFEE FUNCTIONAL GENOME DATABASE (*Coffea sp.*)

### Abstract:

The expression of wrky transcription factor genes in the coffee functional genome was analyzed. It was found 227 reads related to WRKY proteins in *Coffea Arabica*, 205 reads in *Coffea canephora* and 10 reads in *Coffea racemosa*. Among the homologue protein sequences it was also observed some related proteins which have NBS-LRR domain in their structure or probable resistance proteins from different species. The obtained sequences were correlated to different physiological situations specially related to development, abiotic and biotic estresses.

Key words: Functional Genome, Coffee, WRKY

## Introdução

Os fatores de transcrição do tipo WRKY (wrky-tf) são proteínas encontradas em plantas, protistas e algas. Estas proteínas parecem constituir uma complexa superfamília gênica com um ancestral muito antigo (Wu et al, 2005). São caracterizadas por um domínio conservado na seqüência protéica do tipo WRKYGQK (Eulgem et al, 1999), seguido por uma região de *zinc finger* (C2H2 ou C2HC) (Wu et al, 2005). O domínio conservado wrky liga-se a uma seqüência conservada no DNA chamada caixa w (TTGACC/T) (Ulkner & Somssich, 2004) e, esta seqüência pode ser observada em alguns promotores gênicos. As proteínas do tipo WRKY já foram associadas a diversas situações fisiológicas, incluindo o desenvolvimento e a resposta de defesa de plantas a agentes bióticos e abióticos (Ulkner & Somssich, 2004). Com respeito aos estresses produzidos por agentes bióticos ressaltamos os bacterianos e os fúngicos e, no caso dos agentes abióticos, aqueles provocados principalmente por lesão ou estresse hídrico. Também algumas proteínas WRKY parecem estar envolvidas a mecanismos de indução de resistência sistêmica adquirida (SAR). Estas proteínas foram, ainda, observadas como reguladores positivos ou negativos da expressão de genes em diferentes respostas de defesa de plantas (Eulgem, 2006).

Este trabalho analisou a expressão de seqüências do tipo wrky no genoma funcional do café (*Coffea sp.*) (Coffee Genome Project – CAFEST, Brazil). Foram observados 97, 42 e 8 prováveis seqüências gênicas relacionadas a wrky nos bancos *C. arabica*, *C. canephora* e *C. racemosa*, respectivamente.

## Material e Métodos

Este estudo foi realizado utilizando o banco de dados do Projeto Genoma do Café conduzido no Brasil (CAFEST) (<http://www.lge.ibi.unicamp.br/cafe/>). A construção das bibliotecas de cDNAs, a metodologia do sequenciamento e o procedimento para a clusterização das ESTs em *contigs* utilizando a ferramenta do programa *Geneprojects* foram realizadas como descrito em Pereira et al (2004) e Vieira et al, (2006). Neste banco de dados é possível acessar informações de genes expressos em diferentes situações fisiológicas (Tabela 3) de diferentes espécies de café como o *C. arabica*, *C.canephora* e *C. racemosa*.

As reads homólogas às seqüências de proteínas WRKY foram obtidas através da comparação do tipo *tblastn* com seqüências já descritas e depositadas no banco de proteínas do NCBI (NCBI – <http://www.ncbi.nlm.nih.gov/index.html>) utilizando o programa BLAST (Altschul et al, 1990) com um filtro de e-value igual a 10<sup>-04</sup>. As seqüências protéicas homólogas a proteínas WRKY utilizadas para selecionar no banco foram de *Arabidopsis thaliana* (X92976), *Avena fatua* (CAA88331), *Nicotiana tabacum* (BAA7738), *Oryza sativa* (76 em total sendo 31 do grupo 1, 23 do grupo 2 e 22 do

grupo 3 como descritas por Wu et al, 2005), *Petroselinum crispum* ([AF204926](#); [AF121353](#)) e *Solanum tuberosum* ([CAB97004](#); [AAU50687](#)).

## Resultados e Discussão

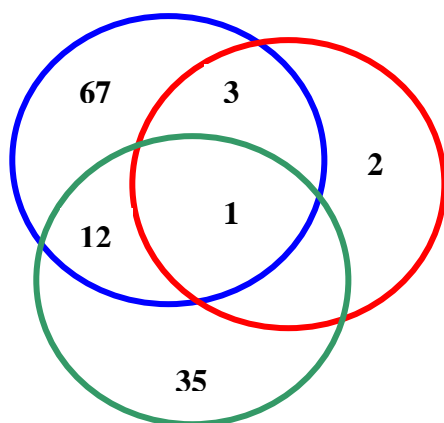
Os dados apresentados na **Tabela 1** mostram o número de *reads* referentes a genes homólogos a fatores de transcrição do tipo *wrky*, encontrados nos bancos de ESTs de *C.arabica*, *C. canephora* e *C. racemosa*. Estas *reads* equivalem a cerca de 83 homologias gênicas diferentes em *C. arabica*, 48 em *C. canephora* e 6 em *C. racemosa*. Tendo em vista que os bancos de *C. arabica*, *C. canephora* e *C. racemosa* estão representados por 47, 9 e 2 bibliotecas de ESTs, respectivamente, é surpreendente o número de seqüências encontradas no banco de *C. canephora*. Dentre as seqüências que apresentam homologia com genes do tipo *wrky* estão genes de proteínas que apresentam região conservada do tipo NBS-LRR, prováveis proteínas de resistência ou a proteína SPF1. A homologia com proteínas de resistência pode ser devido ao domínio *zinc finger*, também presente nas proteínas WRKY. Estas homologias foram observadas nos três bancos de cDNAs estudados.

**Tabela 1** – Número de *reads* e homologias presentes nos bancos de *C. arábica* (CA), *C. canephora* (CC) e *C. racemosa* (CR) em relação a fatores de transcrição tipo *wrky*.

Espécie	Número de Reads	Número aproximado de homologias diferentes	Número de Bibliotecas do banco
CA	227	~ 83	47
CC	205	~ 48	9
CR	10	~ 6	2

Na **Figura 1** pode ser observado que existe apenas uma seqüência comum nos três bancos que refere-se a seqüência da proteína WRKY 30 de *Vitis aestivalis* ([AAR92477](#)). Entre o banco de *C.arabica* e *C.canephora* foi observado 12 homologias gênicas iguais e, entre o banco de *C. arabica* e *C. racemosa* 3 homologias gênicas iguais.

A variação do e-value das *reads* no banco de *C. arabica* é de 0 (83,2%) a  $1 e^{-05}$  (60,2%). A seqüência com e-value igual a 0 equivale a uma proteína tipo WRKY que se liga a DNA de *Nicotiana tabacum* ([BAB 61053](#)) e a seqüência com e-value igual a  $1 e^{-05}$  equivale a uma SPF1 proteína de batata doce ([S51529](#)). No banco de *C. canephora* a variação do e-value vai de  $8e^{-73}$  (49,9%) a 0,0002 (5,86%). O e-value igual a  $8e^{-73}$  equivale a uma proteína putativa de WRKY tipo 4 de *V.aestivalis* ([AAR37421](#)). Esta é a seqüência mais freqüente no banco com 124 *reads*. O e-value igual a 0,0002 (5,86% de similaridade) equivale a uma proteína que se liga a DNA do tipo WRKY 1 de *Parsley* ([S72443](#)). No banco de *C. racemosa* a variação do e-value vai de  $1 e^{-51}$  (58,9%), uma proteína tipo fator de transcrição WRKY 30 de *V. aestivalis* ([AAR92477](#)) a  $6 e^{-15}$  (37%) que representa uma proteína candidata ao gene de resistência RPP8 de *Arabidopsis thaliana*. No banco de *C. arabica* foram observadas 18 seqüências do tipo *no hit* enquanto que no banco de *C.canephora* apenas uma e no banco de *C. racemosa* nenhuma.



**Figura 1** – Intercessão de homologias entre os três diferentes bancos de dados do café. Azul - *C. arábica*; Vermelho - *C. racemosa*; Verde - *C. canephora*.

Com relação aos *contigs* formados foram observados 46, 16 e 2 nos bancos de *C.arabica*, *C.canephora* e *C.racemosa* respectivamente. Além de *contigs* também *singlets*: 51, 26 e 6 respectivamente (**Tabela 2**). A variação do e-value entre os *contigs* do banco de *C. arabica* foi de 0, com 7 representantes (C1, C5, C6, C11, C22, C28, C34) a  $8 e^{-26}$  (C41). Os *contigs* cujos e-value equivalem a zero representam as seqüências das proteínas de ligação a DNA tipo WRKY de *N.tabacum* ([BAB61053](#)), proteína de resistência AAA ATPase ([ABE81238](#)), proteína que se liga a DNA tipo WRKY3

de *N. tabacum* ([BAA77358](#)), provável proteína tipo WRKY 4 de *V. aestivalis* ([AAR37421](#)), proteína fator de transcrição NtEIG-D48 de *N. tabacum* ([BAB16432](#)), proteína NBS-LRR do tipo de resistência ([ABF81448](#)) e a proteína SPF1 de *Ipomoea batatas* ([BAA06278](#)). Esta última com o maior *score*. A sequência com e-value igual a  $8 \times 10^{-26}$  corresponde ao fator de transcrição WRKY 31 de *A. thaliana* ([NP\\_567644](#)).

**Tabela 2** – Número de *Contigs* e *Singlets* formados a partir de *reads* homólogas a fatores de transcrição *wrky* nos bancos de *C. arábica* (CA), *C. canephora* (CC) e *C. racemosa* (CR) e a variação do e-value.

Espécie	Número de <i>Contigs</i>	Variação do e-value nos <i>contigs</i>	Variação do tamanho dos <i>contigs</i> (Kb)	Número de <i>Singlets</i>	Variação do e-value nos <i>singlets</i>
CA	46	0 a $8 \times 10^{-26}$	0,62 a 2,07	51	$2 \times 10^{-81}$ a $7 \times 10^{-07}$
CC	16	0 a $4 \times 10^{-04}$	0,62 a 1,95	26	$3 \times 10^{-65}$ a $2 \times 10^{-04}$
CR	2	$3 \times 10^{-41}$ a $4 \times 10^{-41}$	0,91 a 0,94	6	$1 \times 10^{-51}$ a $6 \times 10^{-15}$

No banco de *C. canephora* a variação do e-value é de 0 com três representantes (C3, C9, C16) a  $4 \times 10^{-04}$  (C8). Os que apresentam e-value igual a zero correspondem a uma proteína putativa do tipo WRKY ([AAP85545](#)), a proteína WRKY 4 de *N. tabacum* ([BAA8603](#)) e a proteína WRKY 31 de *A. thaliana* ([NP\\_567644](#)). A sequência com e-value igual a  $4 \times 10^{-04}$  equivale a uma proteína de *Oryza sativa*.

No banco de *C. racemosa* a diferença do menor e maior e-value é pequena ( $3 \times 10^{-41}$  e  $4 \times 10^{-41}$ ) (C2, C1) e o e-value igual a  $3 \times 10^{-41}$  equivale a uma proteína NBS-LRR do tipo de resistência ([ABF81442](#)) e, o igual a  $4 \times 10^{-41}$  a uma proteína equivalente ao fator de transcrição WRKY 30 de *V. aestivalis* ([AAR92477](#)).

A variação de *singlets* no banco de *C. arábica* vai de  $2 \times 10^{-81}$  (S44 – proteína tipo de resistência de *C. arábica* ([CAC82602](#))) a  $7 \times 10^{-07}$  (S51 – proteína da família WRKY de *A. thaliana* ([NP\\_568777](#))). No banco de *C. canephora* a variação dos *singlets* vai de  $3 \times 10^{-65}$  – S2, proteína SPF1) ([BAA06278](#)) a  $2 \times 10^{-04}$  – S7, proteína WRKY 1 de *Parsley*) ([S72443](#)). E, a variação dos *singlets* no banco do *C. racemosa* vai de  $1 \times 10^{-51}$  (S5 – proteína putativa WRKY 30 de *V. aestivalis*) ([AAR92477](#)) a  $6 \times 10^{-15}$  (S3 – proteína candidata a genes de resistência tipo RPP8 de *A. thaliana*).

É conveniente ressaltar que alguns *contigs* nos bancos de *C. arábica* e *C. canephora* que equivalem a genes *wrky* podem estar completos uma vez que a faixa de comprimento do RNA-m para estas proteínas vai de 1 a 2,3 Kb (**Tabela 2**). Este fato é muito importante para trabalhos futuros. Igualmente seria importante analisar a possível relação das proteínas WRKY do café com os diferentes grupos como já foi realizado em *A. thaliana* e *O. sativa* (Wu et al, 2005).

Tendo em vista a expressão em diferentes sistemas fisiológicos (**Tabela 3**) foi observado que no banco de *C. arábica* ocorrem as maiores expressões, principalmente nas bibliotecas submetidas a estresse por patógenos (ramos infectados por *Xylella* – 9,9% das *reads*) e durante o desenvolvimento de calo (8,8% das *reads*). Ambas as situações confirmam dados na literatura referentes à expressão de proteínas WRKY. No banco de *C. canephora* a maior expressão foi na biblioteca referente a situação fisiológica de sementes precoces com 18 semanas após a polinização com 142 *reads* (71,2%). E, no banco de *C. racemosa* 100 % das *reads* encontradas estão presentes durante o desenvolvimento de frutos.

**Tabela 3** – Relação bibliotecas, ESTs e expressão de fatores de transcrição *wrky* por espécie de *Coffea* sp.

Biblioteca / descrição	Número de <i>Reads</i> encontradas	Percentual
<b><i>C. arábica</i></b>		
Folhas com tratamento de ac. Araquidônico	0	0
Bion positive – plântulas inteiras + células	14	6,2
Calo	20	8,8
Células em suspensão com Bion e brassinosteróides	19	8,4
Colo com e sem Bion – curta	7	3,1
Células em suspensão com sais	5	2,2
Calos embriogênicos	13	5,7
Embrião de frutos	0	0
Embrião de sementes germinando	0	0
Botão floral estádios 1 e 2 – longa	8	3,5
Botão floral estádios 1 e 2 – curta	10	4,4
Botão floral estádios 3 e 4 – curta	4	1,2
Botão floral 6, chumbinho 1, frutos estádios 1 e 2 – longa	19	8,4
Botão floral 6, chumbinho 1, frutos estádios 1 e 2 – curta	19	8,4
Linhagem embriogênica (folha) com indução de 2,4 D	7	3,1
Linhagem embriogênica (folha)	1	0,4

sem indução de 2,4 D		
Plântulas com tratamento de ac. Araquidônico	0	0
Folhas de ponteiro ortotrópicas sem bion - longa	2	0,8
Folhas de ponteiro ortotrópicas sem bion – curta	10	4,4
Folhas plagiotrópicas de plantas adultas sem bion - longa	16	6,6
Folhas plagiotrópicas de plantas adultas sem bion – curta	1	0,4
Raízes com nematóide	0	0
Linhagem embriogênica (calos primários)	4	1,8
Linhagem não embriogênica (folhas) com indução de 2,4 D	3	1,3
Folhas com ferrugem e bicho mineiro	2	0,8
Raízes sem bion – longa	0	0
Raízes com bion – longa	0	0
Raiz e células em suspensão na presença de alumínio - curta	0	0
Ramos infectados com Xylella	21	9,3
Estresse hídrico no campo (pool de tecidos)	6	2,6
Sementes germinando (inteira)	6	2,6
Condições normais (irrigado) (pool de tecidos)	0	0
BP2	1	0,4
FR4	2	0,8
Total	227	100
Biblioteca / descrição	Número de <i>Reads</i> encontradas	Percentual
	<b><i>C. canephora</i></b>	
Calos embriogênicos	15	7,3
Folhas de CC tolerante a seca (estresse hídrico – curta)	0	0
Folhas de CC tolerante a seca – regime de estresse hídrico	10	4,9
SE2 – Semente precoce com 18 meses após a polinização	146	71,2
LF1 - folha	9	4,4
PP1 - pericarpo	14	6,8
SE3 – sementes de idade média com 30 semanas após a polinização	4	2
SE4 – sementes tardias com 42 e 46 semanas após a polinização	7	3,4
Total	205	100
Biblioteca / descrição	Número de <i>Reads</i> encontradas	Percentual
	<b><i>C. racemosa</i></b>	
Fruto	0	0
Fruto verde estádios 1, 2, 3 – curta	0	0
Fruto verde estádios 1, 2, 3	10	100
Total	10	100

## Conclusões

Os dados aqui apresentados demonstram que existe expressão de genes *wrky* no genoma funcional de café. Estes genes parecem ter relação com proteínas de resistência ou que apresentam sítio NBS-LRR em sua estrutura e, expressam-se em diferentes situações fisiológicas como desenvolvimento, estresse por patógenos ou hídrico. São necessários ainda estudos de expressão *in silico* das seqüências do banco CAFEST relacionadas a *wrky* e a validação experimental de sua função biológica.

## Referências Bibliográficas

- .Altschul S., Madden T., Schaffer A., Zhang J., Zhang Z., Mille W., Lipman D.J. (1997) Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res* 25: 3389-3402
- Eugelm, T.; Rushton, P..J.; Schmelzer, E.; Hahlbrock, K.. & Somssich, I.E. (1999) Early nuclear events in plant defence signalling:rapid gene activation by WRKY transcription factors. *The EMBO Journal*, 18 (17):4689-4699.
- Eulgem, T. Dissecting the WRKY Web of Plant Defense Regulators. (2006) [www.plospathogens.org](http://www.plospathogens.org).November , 2: 128-129.
- Pereira, G. A. G. ; Carazzolle, M. F. ; Formighieri, E. F. ; Digiampietri, L ; Arajo, M R . (2004) *Registro INPI no. 005, de 09/01/2004, GeneProjects.*
- Ulkner, B. & Somssich, I.E. (2004) WRKY transcription factors: from DNA binding towards biological function. *Curr.Opin. Plant Biol.*, 7:491-498.
- Vieira, L. G. E., Andrade, A. C., Colombo, C. A., Araújo, A. H., Mehta, A. et al. (2006). Brazilian coffee genome project: an EST-based genomic resource. *Brazilian Journal of Plant Physiology* 18:95-108.
- Wu, K-L; Guo, Z-J; Wang, H-H & Li, J. (2005) The WRKY Family of Transcription Factors in Rice and Arabidopsis and Their Origins. *DNA Research*, 12, 9-26.