

# IDENTIFICAÇÃO E CARACTERIZAÇÃO MOLECULAR DE GENES ENVOLVIDOS NO METABOLISMO DE DITERPENOS ESPECÍFICOS DO CAFEIEIRO

Lucia P. FERREIRA<sup>1</sup>, E-mail: ferreiralp@hotmail.com; Laure PLENER<sup>2</sup>; Pierre Marraccini<sup>3,5</sup>; Luiz Filipe P. PEREIRA<sup>1,3</sup>; Luiz Gonzaga E. VIEIRA<sup>1</sup>; David POT<sup>1,4</sup>

<sup>1</sup>Instituto Agronômico do Paraná, Londrina, PR, <sup>2</sup>Institut National des Sciences Appliquées, Toulouse, France, <sup>3</sup>Embrapa Recursos Genéticos e Biotecnologia, Brasília, DF, <sup>4</sup>CIRAD, UMR DAP Montpellier, France, <sup>5</sup> Embrapa Café, Londrina, PR.

## Resumo:

Apesar de pouco estudado, os lipídeos do café desempenham papel importante na qualidade da bebida e do aroma. Cerca de dez a vinte por cento da fração lipídica do café corresponde a diterpenos que além da provável relação com a qualidade também estão relacionados com a questão do café e a saúde. As espécies de *Coffea* são as únicas capazes de sintetizar os diterpenos cafestol, caveol e seus compostos derivados (16-O-metilcafestol, 16-O-metilcaveol). Visando um entendimento maior dos genes envolvidos na formação dos diterpenos de café, foi realizado um estudo *in silico* da expressão de três genes que codificam para as primeiras etapas da sua via de biossíntese – copalil difosfato sintase (CPS), caureno sintase (KS) e caureno oxidase (KO). Posteriormente foi feita uma análise da expressão destes genes durante o desenvolvimento de frutos de *Coffea arabica* cv. IAPAR 59 e *Coffea canephora* cv. Apoatã. Este estudo permitiu a identificação de um gene para a CPS e KS e dois para a KO. A análise *in silico* revelou níveis e perfis de expressão específicos para cada gene. Os resultados obtidos *in vivo* apresentaram algumas diferenças dos dados obtidos *in silico*, e a comparação de *C. arabica* e *C. canephora* também revelou perfis de expressão diferentes para os genes estudados. Esses dados serão confirmados por RT-PCR e um estudo geral do transcriptoma, usando macroarranjos desenvolvidos a partir do Projeto Genoma Café, será iniciado para identificar os outros genes envolvidos nessa via metabólica.

Palavras-chave: cafestol, caveol, diterpenos, expressão gênica, *Coffea*, maturação.

## MOLECULAR IDENTIFICATION AND CHARACTERIZATION OF GENES INVOLVED IN THE METABOLISM OF COFFEE DITERPENS

### Abstract:

The coffee lipids are important components of aromatic and flavor characteristics of the coffee cup. In the lipid fraction, the diterpens cafestol, kahweol and derived compounds (16-O-methylcafestol, 16-O-methylkahweol) are diterpens specific *Coffea* spp. compounds. Studies have revealed their influence on human health, showing effect on cholesterol level and their action as chemoprotector against toxins with carcinogenic action. In order to increase our knowledge about the genes involved on the coffee diterpens biosynthesis a *in silico* and *in vivo* study of the genes involved in the first stages of their pathway (i.e copalyl diphosphate synthase (CPS), kaurene synthase (KS) and kaurene oxidase (KO)) was performed. For gene expression analysis fruits of *Coffea arabica* cv. IAPAR 59 and *Coffea canephora* cv. Apoatã at different stages of maturation was used. *In silico* studies allowed the identification of one gene for CPS and KS and two for KO. It also revealed different expression patterns specific to each gene in terms of level and timing of expression. Results of expression *in vivo* were slightly different from the ones obtained *in silico*, and the comparison of *C. canephora* and *C. arabica* also revealed some differences. These data will be confirmed by RT-PCR and a whole transcriptome approach using microarrays based on the Genoma Café Project will be initiated to identify other genes involved in this biosynthesis pathway.

Key words: cafestol, kahweol, diterpens, gene expression, *Coffea*, maturation.

### Introdução

Recentemente, foram finalizados dois projetos de sequenciamento em larga escala de genes expressos EST de café: o projeto “Genoma Café” (Vieira *et al.* 2006: em torno de 153.739 ESTs) e o projeto Nestlé/Cornell (Lin *et al.* 2005: 46.914 ESTs). Estes projetos visam fornecer novos conhecimentos sobre as bases genéticas e moleculares de características agronômicas importantes no cafeeiro. O fato das seqüências ESTs de café estarem agora disponíveis abre novos caminhos para a identificação dos principais genes que controlam as vias biológicas importantes, como as implicadas na biossíntese dos diterpenos específicos do café, o cafestol e o caveol.

Espécies de *Coffea* são as únicas capazes de sintetizar os diterpenos cafestol, caveol e seus compostos derivados (16-O-metilcafestol, 16-O-metilcaveol). São conhecidos alguns efeitos destes diterpenos de café na saúde humana, agindo como substância com capacidade de elevar o nível de colesterol, como quimioprotetor contra toxinas hepatocarcinogênicas e como moléculas com ação antioxidante. Muitos usos para o cafestol e caveol já foram patenteados. O óleo de café, que contém estes dois diterpenos, foi citado como apropriado para uso como protetor solar. Em combinação com substâncias cosméticas para suporte, formulações para uso tópico contendo quantidades efetivas de cafestol foram patenteadas para prevenção ou tratamento de pele. O uso de palmitatos de cafestol e caveol para tratamento de lesões com potencial cancerígeno.

Até o momento, não há nenhum relato sobre os genes envolvidos no metabolismo destes diterpenos específicos do café, particularmente durante a fase de desenvolvimento dos grãos, mas também em outros tecidos, como nas folhas, onde eles são também encontrados. Assim, aplicações potencialmente importantes dos diterpenos de café em cosméticos e na indústria farmacêutica, fazem com que a análise molecular dessas vias de biossíntese mereça especial atenção para agregar valor ao café e seus subprodutos.

Portanto, este trabalho propõe aumentar nosso conhecimento sobre o metabolismo de diterpenos em café através da identificação e caracterização dos genes envolvidos na biossíntese de cafestol e de caveol, analisando a expressão destes genes em vários tecidos de frutos em diferentes estádios de desenvolvimento. Estas informações poderão ser usadas para definir o efeito da regulação da expressão destas seqüências na concentração dos diterpenos de café.

A estratégia seguida nesse estudo foi: i) análise *in silico* dos dados ESTs disponíveis para avaliar o número de genes que codificam para cada enzima e ter uma idéia de seus perfis de expressão e ii) análise *in vivo* destes genes durante o desenvolvimento do grão em duas espécies de *Coffea* (*C. arabica* cv. IAPAR 59 e *C. canephora* cv Apotã).

## Material e Métodos

### 1. Estudo *in silico*

Estudo *in silico* da expressão dos genes copalyl diphosphate synthase (CPS), kaurene synthase (KS) and kaurene oxidase foi realizado a partir de 37 bibliotecas EST (Expressed Sequence Tags) geradas pelo Projeto Genoma Café. Como estes ESTs são agrupamentos de diferentes bibliotecas, obtidas, principalmente, a partir de *C. arabica*, e correspondentes a diferentes tipos celulares e tecidos em diferentes estádios de desenvolvimento, esta análise não teve como único objetivo determinar o perfil de expressão cada um dos genes, mas também verificar se um gene apresenta um perfil de expressão diferente dos outros e ainda ter uma idéia do nível de expressão cada um dos genes estudados.

Análise *in silico* também foi realizada a partir de seqüências Nestlé/Cornell disponíveis na Internet (<http://harvest.ucr.edu>). Este banco contém quatro bibliotecas de ESTs de frutos de *C. canephora*, em diferentes estádios de desenvolvimento e de diferentes tecidos, o que permitiu estudar a expressão destes genes durante o desenvolvimento do grão.

Além disso, uma avaliação da variabilidade de expressão destes genes durante o desenvolvimento do grão foi feito utilizando o banco de dados de Nestlé. A partir do número de cópias de cada gene nas bibliotecas, do número total de ESTs nas bibliotecas e através da ajuda de testes estatísticos, foi realizado um estudo para detectar expressões diferenciais entre as bibliotecas com o objetivo de compreender onde e como são expressos os genes estudados. Para detectar estas expressões diferenciais foi utilizado Programa IDEG6 (<http://telethon.bio.unipd.it/bioinfo/IDEG6>).

### 2. Estudo *in vivo*

Foram utilizados frutos de *C. arabica* cv. IAPAR 59 e de *C. canephora* cv. Apotã coletados a cada quatro semanas desde a floração até a completa maturação, durante a safra 2005/2006. Tecidos do fruto (perisperma, endosperma, polpa) foram separados e usados independentemente para extração do RNA total (Rogers *et al.*, 1999).

Sondas internas foram amplificadas por PCR usando cDNA frutos inteiros de *C. arabica* e de *C. canephora*, sob uma mesma concentração, como molde, oligonucleotídeos específicos para cada enzima de interesse (CPS, KO e KS) e *Taq* DNA polimerase. Estas sondas foram purificadas por precipitação com 10% (v/v) NaAc pH 5,2; 3X (v/v) etanol 100%, ressuspendidas em água e quantificadas.

Para os Northern blots foram usados 10 µg de RNA em gel de agarose 1,4%. O protocolo utilizado para as análises é o mesmo descrito por Geromel *et al.* (2006).

## Resultados e Discussão

### 1. Estudo *in silico*

Para CPS, 63 seqüências foram identificadas no Projeto Genoma Café, incluindo 44 de *C. arabica* e 19 de *C. canephora*. Outras cinco seqüências foram identificadas na biblioteca de cDNA de *C. canephora* desenvolvida pela Nestlé/Cornell. Estas 68 seqüências formaram um *contig* de cDNA completo (2.671 pb), correspondendo a uma proteína de 825 aa.

Para KS foram identificadas 26 seqüências no Projeto Genoma Café. Outras cinco foram identificadas em bibliotecas de cDNA desenvolvida pela Nestlé/Cornell. Estas 31 seqüências formaram um *contig* completo (2.700 pb), que corresponde a uma proteína de 788 aa.

Para KO foram identificadas 82 seqüências no Projeto Genomacafé, sendo 80 de *C. arabica* e duas de *C. canephora*. Outras 10 seqüências foram identificadas na biblioteca de cDNA desenvolvida pela Nestlé/Cornell. Estas 92 seqüências formaram dois *contigs* de 82 (KO1) e 10 seqüências (KO2). O *contig* que contém o maior número de seqüências (KO1) apresenta cDNA completo com 2.283 pb, o que corresponde a uma proteína de 517 aa. O segundo *contig* (KO2) corresponde a uma proteína incompleta na qual falta ao torno de 40 aa na região C-terminal. Estas duas proteínas (KO1 e KO2) apresentam uma identidade de 86% e uma homologia de 93%.

Foi possível fazer uma análise de expressão dos genes durante o desenvolvimento do grão usando os dados de Nestlé/Cornell, pois esta biblioteca apresenta frutos em três estádios distintos de maturação (início, meio e fim). Esta análise indicou diferenças de expressão ao longo do desenvolvimento do grão para CPS e KO, mas não para KS. Observa-se assim que KO é fortemente expressa no fruto inteiro jovem (oito ESTs expressos em frutos jovens, seguido de uma fraca expressão no meio e no final da maturação, respectivamente 0 e 1 EST). A tendência parece oposta para CPS, com uma maior expressão no final da maturação (cinco seqüências expressas). Não houve nenhuma expressão nas fases mais jovens de desenvolvimento do fruto. Na polpa foi observada uma expressão fraca para CPS e KS e uma expressão um pouco maior de KO.

Já foi relatado que a enzima KS tem uma alta atividade, o que pode explicar o baixo nível de transcritos detectados em comparação com os outros genes. Kawaide *et al.* (2000) e Fleet *et al.* (2003) também observaram que uma subexpressão gênica de KS não tem efeito pronunciado na síntese de ent-kaurene. A falta de expressão diferencial observada *in silico* durante o desenvolvimento do grão de café suporta a hipótese que a KS não é a enzima limitante da via da biosíntese do ent-Kaurene, o precursor dos diterpenos.

Também, já foi observado que a KO não é a enzima limitante na via da biosíntese dos diterpenos. (Swain *et al.* 2005). A baixa expressão do gene que codifica para esta enzima em *Arabidopsis* não induziu aumento da concentração de ácidos giberélicos e nem mesmo ácido ent-kaurenóico. Como foi observada uma expressão diferencial de KO *in silico* durante o desenvolvimento do fruto, será importante avaliar a atividade enzimática para verificar o papel desta enzima nesta via metabólica.

A análise *in silico* revelou que o gene CPS apresenta uma expressão diferencial, considerando a análise de diferentes tecidos e também ao longo da maturação. A CPS é a primeira enzima da via, o que lhe confere um papel evidente como regulador no controle da entrada de metabólitos. Por outro lado, esta enzima está sujeita a uma forte regulação e a uma retroinibição pelos ácidos giberélicos, sendo, portanto esta enzima a que controla a biosíntese das giberelinas bioativas (Fleet *et al.* 2003). Os resultados observados *in silico* concordam com esse papel de regulação.

## 2. Estudo *in vivo*

- *C. arabica*

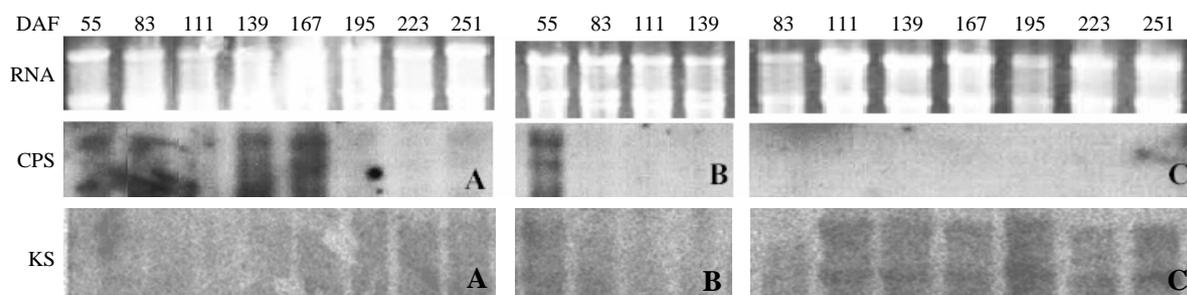
As hibridizações realizadas permitem determinar que a expressão de CPS ocorre na polpa (139–167 Dias Após a Floração – DAF) e no perisperma (55 DAF) (Figura 1). Não foi detectada expressão em endosperma.

Para o gene KS a expressão na polpa é observada em todo ciclo de maturação (195 – 223 DAF). Para o perisperma a expressão é detectada no início do desenvolvimento do fruto (55 DAF) e em endosperma há uma expressão constante e fraca durante todo o desenvolvimento do fruto (Figura 1).

Os sinais de hibridização para KO1 foram muito fracos. A expressão do gene em polpa ocorre de 83 a 167 DAF. Em perisperma, os dados são os mesmos observados para os genes de CPS e KS (55 DAF). Para o endosperma há uma expressão no início e no meio do desenvolvimento (83 – 139 DAF) (dado não mostrado).

Certos terpenos têm função bem caracterizada no crescimento e no desenvolvimento vegetal. Por exemplo, as giberelinas, um importante grupo de fitorreguladores, são diterpenos (Taiz e Zeiger, 2004). A expressões dos três genes da via de diterpenos observadas em perisperma de *C. arabica* ocorrem ao mesmo tempo (55 DAF) em que há o crescimento do fruto e o desenvolvimento deste tecido.

Os carotenóides de cores vermelha, amarela e laranja são tetraterpenos que agem como pigmentos acessórios na fotossíntese e protegem os tecidos fotossintéticos contra a fotoxidação. O fitorregulador ácido abscísico é um terpeno C<sub>15</sub>, produzido pela degradação de um precursor carotenóide (Taiz e Zeiger, 2004). Nesta fase do estudo, não se pode afirmar, ainda, se um aumento na expressão dos genes CPS, KO e KS corresponde a um aumento na atividade da via biossintética de diterpenos ou de ácidos giberélicos, que estão diretamente implicados na maturação do fruto. Entretanto, em polpa de *C. arabica* as expressões dos genes estudados são mais intensas na fase de maturação, com a aparição da coloração vermelha dos frutos.



**Figure 1:** Gel de Northern e hibridizações de RNAs de *C. arabica* com as sondas CPS e KS. **A.** RNA de polpa; **B** RNA de perisperma; **C.** RNA de endosperma.

- *C. canephora*

Embora a expressão de transcritos de CPS no perisperma seja claramente detectada em *C. arabica*, ela não foi observada no mesmo tecido de *C. canephora*. Este fato surpreende, já que o perisperma é um tecido muito ativo no café e sua evolução é muito importante ao longo da maturação do fruto. Há um sinal fraco para polpa que não permite determinar uma expressão diferencial ao longo do ciclo de maturação. Assim como para *C. arabica*, nenhuma expressão foi detectada no endosperma (dado não mostrado).

Os resultados de Northern blot obtidos para o gene KO foram muito fracos. Não foi possível identificar expressão para polpa e perisperma. A expressão mais forte é observada em endosperma, na fase final da maturação (167 – 223 DAF).

Para KS, não há resultados de expressão conclusivos para esta espécie (dado não mostrado).

Uma comparação entre os resultados obtidos através das diferentes análises (*in silico* e *in vivo*) indicou que em *C. arabica*, uma alta expressão de KO1 foi observada *in silico* (71 seq.), mas este resultado não foi observado *in vivo*. Para o gene CPS houve uma alta expressão *in vivo* e uma média expressão *in silico* (44 seq.). Para KS foi possível uma expressão após a hibridização, muito embora haja poucas seqüências (36) nos bancos de dados.

Nos dados Nestlé/Cornell, o número de ESTs KO aposentam uma forte expressão *in silico* nos frutos jovens, o que difere da fraca expressão observada no fim do desenvolvimento de endosperma.

## Conclusões

Os resultados obtidos *in vivo* serão confirmados usando PCR em tempo real com o objetivo de solucionar problemas de possíveis inespecificidades das sondas que podem complicar a interpretação dos Northern-blots (especialmente no caso do KO1 e KO2). Além de *C. arabica* e *C. canephora*, o *C. eugenoides* será estudado para tentar de entender as diferenças que parecem existir entre as duas espécies de interesse comercial. Adicionalmente à análise específica desses três genes, uma estratégia mais ampla usando os macroarranjos desenvolvidos a partir dos dados do Genoma Café será usada para identificar outros genes da via da biossíntese dos diterpenos que ainda não foram identificados.

## Agradecimentos

Apoio financeiro do Consórcio Brasileiro de Pesquisa e Desenvolvimento do Café (CBP&D-Café) e do CNPq Proj 470875/2004-06. Lucia P. Ferreira - Bolsista CBP&D-Café.

## Referências Bibliográficas

- Fleet CM, Yamaguchi S, Hanada A, Kawaide H, David CJ, Kamiya Y, Sun TP (2003). Overexpression of *AtCPS* et *AtKS* in *Arabidopsis* confers increased *ent*-kaurene production but no increase in bioactive gibberellins. *Plant Physiol. Vol 13* 830-839.
- Geromel C., Ferreira L.P., Guerreiro S.M.C., Cavallari A.A., Pot D., Pereira L.F.P., Leroy T., Vieira L.G.E., Mazzafera P., Marraccini P.(2006) Biochemical and genomic analysis of sucrose metabolism during coffee (*Coffea arabica*) fruit development.. *J. Exp. Bot.* 57:3243-3258.
- Kawaide H, Sassa T, Kamiya Y (2000). Functional analysis of the two interacting cyclase domains in entkaurene synthase from the fungus *Phaeosphaeria* sp. L487 and a comparison with cyclases from higher plants. *J. Biol. Chem.* 275(4):2276-80.
- Lin, C.; Mueller, L.A.; Mc Carthy, J.; Crouzillat, D.; Pétiard, V.; Tanksley, S.D. (2005) Coffee and tomato share common gene repertoires as revealed by deep sequencing of seed and cherry transcripts. *Theor Appl Genet.* 112:114-130.
- Rogers W.J., Michaux S., Bastin M., Bucheli P. (1999) Changes to the contents of sugars, alcohols, myo-inositol, carboxylic acids and inorganic anions in developing grains from different varieties of Robusta (*Coffea canephora*) and Arabica (*Coffea arabica*). *Plant Sci* 149:115-123.
- Swain SM, Singh DP, Helliwell CA, Poole AT (2005). Plants with increased expression of *ent*-kaurene oxidase are resistant to chemical inhibitors of this gibberellin biosynthesis enzyme. *Plant Cell Physiol.* 46(2):284-91.
- Taiz L., Zeiger E. (2004) Metabólitos secundários e defesa vegetal. In: *Fisiologia Vegetal* 3ªed. 309-334.
- Vieira, L.G.E.; Andrade, A.C.; Colombo, C.A.; Moraes, A.H.de.A.; Metha, A.; Carvalho de Oliveira, A.; Labate, C.A.; Marino, C.L.; Monteiro-Vitorello, C.de.B.; Monte, D. de C.; Giglioti, E.; Kimura, E.T.; Romano, E.; Kuramae, E.E.; Lemos, E.G.M.; Pereira de Almeida, E.R.; Jorge, E.C.; Albuquerque, E.V.S.; da Silva, F.R.; Vinecky, F.; Sawazaki, H.E.; Dorry, H.F.A.; Carrer, H.; Abreu, I.N.; Batista, J.A.N.; Teixeira, J.B.; Kitajima, J.P.; Xavier, K.G.; Maria de Lima, L.; Aranha de Camargo, L.E.; Pereira, L.F.P.; Coutinho, L.L.; Lemos, M.V.F.; Romano, M.R.; Machado, M.A.; Costa, M.M. do. C.; Grossi, de Sá M.F.; Goldman, M.H.S.; Ferro, M.I.T.; Tinoco, M.L.P.; Oliveira, M.C.; Van Sluys, M-A.; Shimizu, M.M.; Maluf, M.P.; Souza da Eira, M.T.; Guerreiro Filho, O.; Arruda, P.; Mazzafera, P.; Mariani, P.D.S.C.; de Oliveira, R.L.B.C.; Harakava, R.; Balbao, S.F.; Tsai, S.M.; di Mauro, S.M.Z.; Santos, S.N.; Siqueira, W.J.; Costa, G.G.L.; Formighieri, E.F.; Carazzolle, M.F.; Pereira, G.A.G. (2006). Brazilian coffee genome project: an EST-based genomic resource. *Braz. J. Plant Physiol.*, 18(1):95-108.