



UNIVERSIDADE  
ESTADUAL DE LONDRINA

---

SUZANA TIEMI IVAMOTO

**DITERPENOS EM *COFFEA ARABICA*:**  
ASPECTOS BIOQUÍMICOS, E CARACTERIZAÇÃO *IN SILICO*  
E TRANSCRICIONAL DE GENES DE Cyt P450s  
CANDIDATOS NA BIOSÍNTESE DE CAFESTOL E CAVEOL

**SUZANA TIEMI IVAMOTO**

**DITERPENOS EM *COFFEA ARABICA*:**  
**ASPECTOS BIOQUÍMICOS, E CARACTERIZAÇÃO *IN SILICO***  
**E TRANSCRICIONAL DE GENES DE Cyt P450s**  
**CANDIDATOS NA BIOSÍNTESE DE CAFESTOL E CAVEOL**

Dissertação de Mestrado apresentado ao Programa de Pós Graduação em Genética e Biologia Molecular, da Universidade Estadual de Londrina, como requisito parcial para a obtenção do título de Mestre.

Orientador: Dr. Luiz Filipe Protasio Pereira

Londrina  
2012

**Catálogo elaborado pela Divisão de Processos Técnicos da Biblioteca Central da  
Universidade Estadual de Londrina**

**Dados Internacionais de Catalogação-na-Publicação (CIP)**

I93d Ivamoto, Suzana Tiemi.

Diterpenos em *Coffea arabica* : aspectos bioquímicos, e caracterização *in silico* e transcripcional de genes Cyt P450s candidatos na biossíntese de cafestol e caveol / Suzana Tiemi Ivamoto. – Londrina, 2012.  
89 f. : il.

Orientador: Luiz Filipe Protasio Pereira.

Dissertação (Mestrado em Genética e Biologia Molecular) – Universidade Estadual de Londrina, Centro de Ciências Biológicas, Programa de Pós-Graduação em Genética e Biologia Molecular, 2012.

Inclui bibliografia.

1. Café – Melhoramento genético – Teses. 2. Diterpenos – Teses. 3. Genética – Expressão – Teses. 4. Genética vegetal – Teses. I. Pereira, Luiz Filipe Protasio. II. Universidade Estadual de Londrina. Centro de Ciências Biológicas. Programa de Pós-Graduação em Genética e Biologia Molecular. III. Título.

CDU 631.52:633.73

**SUZANA TIEMI IVAMOTO**

**DITERPENOS EM *COFFEA ARABICA*:**  
ASPECTOS BIOQUÍMICOS, E CARACTERIZAÇÃO *IN SILICO* E  
TRANSCRICIONAL DE GENES DE Cyt P450s CANDIDATOS NA  
BIOSSÍNTESE DE CAFESTOL E CAVEOL

Dissertação de Mestrado apresentado ao Programa de Pós Graduação em Genética e Biologia Molecular, da Universidade Estadual de Londrina, como requisito parcial para a obtenção do título de Mestre.

**COMISSÃO EXAMINADORA**

---

Dr. Luiz Filipe Protasio Pereira  
IAPAR - Embrapa Café – Londrina - PR

---

Dr. Hugo Bruno Correa Molinari  
Embrapa Agroenergia – Brasília - DF

---

Dr. Laurival Antonio Vilas Boas  
UEL – Londrina - PR

Londrina, 16 de Fevereiro de 2012.

## **AGRADECIMENTOS**

A Universidade Estadual de Londrina, ao curso de Pós graduação em Genética e Biologia Molecular e seus professores pela oportunidade, estrutura e pelos diferentes canais de aprendizagem oferecidos, os quais foram fundamentais ao longo desses 2 anos.

Ao Instituto Agrônomo do Paraná, pelo acolhimento e estrutura disponibilizada durante a realização deste mestrado.

Ao meu orientador Dr. Luiz Filipe Protasio Pereira, pelos ensinamentos, confiança em mim depositada, orientações para a melhoria deste trabalho e pela atenção dedicada para a realização deste projeto.

Aos doutores Luiz Gonzaga Esteves Vieira, Douglas Silva Domingues e David Pot pela ajuda na realização deste projeto e pelas contribuições recebidas para melhor direcioná-lo.

Aos doutores Hugo Bruno Correa Molinari e Laurival Antonio Vilas Boas pela disposição em avaliar este projeto.

À doutora Maria Brígida dos Santos Scholz e a mestre Cíntia Sorane Good Kitzberger, pelo auxílio nas análises de quantificações por CLAE e estatística.

À minha mãe Maria Sadako Ivamoto, ao meu irmão Eduardo Massaaki Ivamoto e ao meu namorado Angelo Heidi Suzuki, pelo carinho, compreensão, apoio à minha educação e suporte nos momentos menos fáceis.

Aos meus amigos do laboratório de Biotecnologia Vegetal do IAPAR, por todo o auxílio concedido, pela ótima convivência que criamos no nosso ambiente de trabalho e principalmente pela amizade!

Ao Consórcio Pesquisa Café e a Financiadora de Estudos e Projetos (FINEP) pelo financiamento deste projeto de pesquisa.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico pela concessão da bolsa de estudo.

E a todos que de alguma maneira contribuíram direta ou indiretamente para a realização e desenvolvimento deste trabalho. Muito Obrigada!

IVAMOTO, Suzana Tiemi. **Diterpenos em *Coffea arabica***: aspectos bioquímicos, e caracterização *in silico* e transcricional de Cyt P450s candidatos na biossíntese de cafestol e caveol 89 f. Dissertação de Mestrado (Pós Graduação em Genética e Biologia Molecular), Universidade Estadual de Londrina, Londrina, 2012.

## RESUMO

O Brasil é o maior produtor e exportador mundial de café e o segundo maior mercado consumidor. Estudos visando a melhoria da qualidade da bebida e da tolerância de plantas a estresses bióticos e abióticos são importantes para o agronegócio café. As propriedades nutracêuticas da bebida e mecanismos de defesa da planta estão relacionadas aos diterpenos presentes na fração lipídica de grãos de café, como o cafestol (Caf) e caveol (Cav). As etapas finais de biossíntese destes compostos ainda não foram elucidadas, mas estudos relacionados ao metabolismo secundário de plantas indicam o envolvimento de membros da família gênica das citocromos P450s (Cyt P450). No intuito de caracterizar aspectos bioquímicos e genéticos da biossíntese destes compostos, foram realizadas caracterizações bioquímicas de diterpenos em frutos de café e a caracterização molecular de genes da família das P450s em *Coffea spp.* Foram quantificados por CLAE os teores de Caf e Cav em: i) botão floral, flor e perisperma de frutos (30 e 210DAF e ii) frutos com 120DAF tratados com metil jasmonato 2 $\mu$ M (MJ). Há um pico de concentração de Cav em perisperma de frutos aos 120DAF (284% mais caveol do que em 90DAF), com queda acentuada aos 150DAF e 396% mais Caf em 120DAF do que em 150DAF. Há uma inversão nas concentrações de Caf e Cav do botão floral/flor (predomínio de Caf) para os frutos (predomínio de Cav). A concentração de diterpenos em frutos (120DAF) tratados com 2 $\mu$ M metil jasmonato (MJ) foi determinada durante o período de 24, 48 e 96 horas. Frutos após 48 horas do tratamento com MJ apresentaram a maior diferença na concentração de diterpenos em relação aos frutos não tratados com um aumento de 48 e 19% de Caf e Cav, respectivamente. Neste trabalho foram também identificadas 84 Cyt P450s transcricionalmente ativas em *Coffea spp.*, provenientes de sequências ESTs do Projeto Brasileiro Genoma Café. Destes, oito genes candidatos foram escolhidos para avaliação transcricional por qPCR em folhas, flores e frutos em 3 períodos de desenvolvimento (90DAF, 120DAF e 150DAF). Sete genes apresentaram expressão diferencial entre os tecidos analisados, Os contigs C22\_1 e C25\_1 expressaram mais em folhas e apresentaram baixa atividade transcricional nos demais tecidos analisados. Por outro lado os contigs C19\_2, C28\_1 e C37\_2 tiveram uma alta expressão relativa em flor enquanto que os contigs C33\_1 e C107\_1 apresentaram uma expressão maior em frutos com 90DAF. Esses últimos cinco contigs têm grande potencial de estarem envolvidos nas etapas finais da via metabólica de Caf e Cav.

**Palavras chave:** Diterpenos. Cyt P450. Cafestol. Caveol. *Coffea*. qPCR.

**Apoio Financeiro:** Consórcio Pesquisa Café, CNPq e FINEP.

IVAMOTO, Suzana Tiemi. **Diterpenes in *Coffea arabica***: biochemical aspects, in silico and transcriptional characterization of Cyt P450s candidate genes in the biosynthesis of cafestol and kahweol. 89p. Master's Degree – Universidade Estadual de Londrina, Londrina, 2012.

## ABSTRACT

Brazil is coffee largest producer and exporter, and the second largest consumer market. Studies aiming the beverage quality improvement and plant tolerance to biotic and abiotic stresses are important to coffee agrobusiness. Nutraceutical properties in beverage and plant defense mechanisms are directly related to diterpenes present in the lipid fraction of coffee beans, such as cafestol (Caf) and kahweol (Cav). The final steps of Caf and Cav biosynthesis have not been elucidated yet, but plant secondary metabolism studies indicate the involvement of members from cytochrome P450s gene family (Cyt P450). In order to depict genetic and biochemical aspects of Caf and Cav biosynthesis, it was performed the biochemical quantification of these diterpenes in coffee fruits and the molecular characterization of members from Cyt p450's family in *Coffea spp.* Levels of Caf and Cav were quantified by HPLC on: i) bud flower, flower and fruit's perisperm (30 to 210DAF) and ii) at 120 DAF fruits treated with methyl jasmonate 2 $\mu$ M (MJ). The highest concentration of Cav occurred in the perisperm of 120DAF fruits (284% higher than in 90DAF) decreasing at 150DAF and 396% more Caf in 120DAF than 150DAF. There is an inversion in the Caf and Cav concentrations from the bud flower and flower (Caf prevalence) to fruits (predominantly Cav). Diterpene concentration on 120DAF fruits (120DAF) treated with 2 $\mu$ M of methyl jasmonate (MJ) was determined after 24, 48 and 96 hours. After 48 hours, MJ-treated fruits showed an increase of 48 and 19% of Caf and Cav, respectively, when compared to untreated fruits. In this study we also identified 84 Cyt P450s transcriptionally active in *Coffea spp.*, from the Brazilian Coffee Genome Consortium (EST sequences). Eight candidate genes were chosen for further transcriptional analysis (qPCR) in leaves, flowers and fruits at three developmental stages (90DAF, 120DAF and 150DAF). Seven genes displayed differential expression among tissues. The contigs C22\_1 and C25\_1 had preferential expression in leaves and showed low transcriptional activity in other tissues. Moreover the contigs C19\_2, C28\_1 and C37\_2 had a high relative expression in flower while the contigs C33\_1 and C107\_1 showed a major expression in fruits with 90DAF. These last five contigs have a great potential to be involved in the final stages of Caf and Cav metabolic pathway.

**Keywords:** Diterpenes. Cyt P450. Cafestol. Kahweol. *Coffea*.

**Financial support:** Brazilian Coffee Research, CNPq and FINEP.

## LISTA DE TABELAS

<b>Tabela 1</b> - Picos de quantificação dos diterpenos nos tecidos de frutos e folha .....	16
<b>Tabela 2</b> - Teores de cafestol e caveol de acordo com a forma de preparo .....	18
<b>Tabela 3</b> - Classificação das Cyt P450 dos contigs de <i>Coffea</i> spp. ....	27
<b>Tabela 4</b> - Classificação das Cyt P450 ( <i>BlastX</i> - TAIR e NCBI/CDD) .....	28
<b>Tabela 5</b> - Família Gênica das Cyt P450 de <i>Coffea</i> spp. do tipo não A .....	30
<b>Tabela 6</b> - Família Gênica das Cyt P450 de <i>Coffea</i> spp. do tipo A .....	30
<b>Tabela 7</b> - Prováveis funções das Cyt P450 de <i>Coffea</i> .....	32
<b>Tabela 8</b> - Concentrações médias de Caf e Cav ao longo do desenvolvimento do fruto (safra 2010/2011) .....	46
<b>Tabela 9</b> - Concentrações médias de diterpenos em frutos tratados com metiljasmonato (safra 2009/2010) .....	48
<b>Tabela 10</b> - Resultado da Seleção dos Genes Candidatos .....	49
<b>Tabela 11</b> - Lista dos <i>Primers</i> desenhados para os Genes Candidatos .....	51

## LISTA DE FIGURAS

<b>Figura 1</b> - Fórmula estrutural do cafestol e caveol .....	15
<b>Figura 2</b> - Estádios de desenvolvimento do fruto de cafeeiro .....	17
<b>Figura 3</b> - Via MVA e MEP da biossíntese de isoprenóides e os seus respectivos compostos sintetizados .....	20
<b>Figura 4</b> - Envolvimento de Cyt P450 na via de biossíntese de giberelinas .....	22
<b>Figura 5</b> - Hipótese sobre a sequência das etapas finais da biossíntese dos diterpenos Caf e Cav .....	23
<b>Figura 6</b> - Árvore filogenética parcial das Cyt P450 .....	31
<b>Figura 7</b> - Perisperma do fruto de cafeeiro .....	42
<b>Figura 8</b> - Níveis médios de Caf e Cav por CLAE de botão floral, flor e do perisperma de frutos ao longo do seu desenvolvimento .....	46
<b>Figura 9</b> - Níveis médios de Caf e Cav por CLAE nas amostras controle e com aplicação de metil jasmonato (safra de 2009/2010) .....	48
<b>Figura 10</b> - Perfis Transcricionais dos genes candidatos de <i>Coffea arabica</i> Cyt P450 da família tipo A .....	54
<b>Figura 11</b> - Perfis Transcricionais de genes candidatos de <i>C. arabica</i> Cyt p450 da família tipo não A .....	55
<b>Figura 12</b> - Hipótese da via de biossíntese de Caf e Cav com a inclusão dos genes candidatos de Cyt P450s identificados. ....	56

## ABREVIATURAS E SIGLAS

<b>ABA</b>	Ácido abscísico
<b>B</b>	Brassinoesteróis
<b>C3'H:</b>	p-coumaroil-3-hidroxilase
<b>C4H</b>	cinamato-4-hidroxilase
<b>CA</b>	Caroteno
<b>CAF</b>	Cafestol
<b>CAV</b>	Caveol
<b>CCA</b>	Software codon code aligner
<b>CL</b>	Clorofila
<b>CLAE</b>	Cromatografia líquida de alta eficiência
<b>Cyt P450</b>	Citocromo P450
<b>DAF</b>	Dias após florada
<b>DMAPP</b>	Dimetilalil difosfato
<b>DXS</b>	DXP sintase
<b>DXR</b>	DXP reductoisomerase
<b>E</b>	Esteróis
<b>EST</b>	Sequências nucleotídicas expressas
<b>FINEP</b>	Financiadora de estudos e projetos
<b>FPP</b>	Farnesil difosfato
<b>GAPDH</b>	Gliceraldeído-3-fosfato desidrogenase
<b>GPP</b>	Geranil difosfato
<b>GGPP</b>	Geranil geranil difosfato
<b>HMGR</b>	3-hidroxi-metilglutaril-Coenzima A redutase
<b>IDI</b>	Isopentenil isomerase
<b>IDS</b>	Isopentenil/dimetilalil difosfato sintase
<b>IPP</b>	isopentenil difosfato
<b>KO</b>	Caureno oxidase
<b>MEP</b>	Via plastidial ou via não mevalonato da biossíntese de isoprenóides
<b>MVA</b>	Via citosólica ou via do mevalonato da biossíntese de isoprenóides
<b>qPCR</b>	PCR em tempo real
<b>UBQ10</b>	Poliubiquitina 10

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b> .....	<b>11</b>
<b>2</b>	<b>OBJETIVO GERAL</b> .....	<b>12</b>
2.1	OBJETIVOS ESPECÍFICOS .....	12
<b>3</b>	<b>FUNDAMENTAÇÃO TEÓRICA</b> .....	<b>13</b>
3.1	A IMPORTÂNCIA SÓCIO-ECONÔMICA DA CAFEICULTURA .....	13
3.2	Os LIPÍDEOS DO CAFÉ .....	13
3.3	ÁCIDOS GRAXOS DO CAFÉ .....	14
3.4	Os DITERPENOS DO CAFÉ .....	15
3.5	BIOSSÍNTESE DE DITERPENOS .....	18
3.5.1	Síntese de Isoprenóides .....	18
3.5.2	A Super Família das Cyt P450s .....	21
3.5.3	A Hipótese Para a Biossíntese de Cafestol e Caveol .....	22
<b>4</b>	<b>ARTIGO 1: A SUPER FAMÍLIA DAS CITOCROMOS P450S EM COFFEA</b> .....	<b>24</b>
4.1	INTRODUÇÃO .....	24
4.2	MATERIAIS E MÉTODOS .....	25
4.2.1	Identificação das Sequências de Cyt P450 .....	25
4.2.2	Validação e Caracterização das Cyt P450 .....	26
4.2.3	Árvore Filogenética .....	26
4.3	RESULTADOS E DISCUSSÃO .....	26
4.4	REFERENCIAL BIBLIOGRÁFICO .....	34
<b>5</b>	<b>ARTIGO 2 - BIOSÍNTESE DE DITERPENOS EM COFFEA ARABICA: CARACTERIZAÇÃO BIOQUÍMICA E ANÁLISE DO PERFIL TRANSCRICIONAL DE GENES CANDIDATOS</b> .....	<b>40</b>
5.1	INTRODUÇÃO .....	40
5.2	MATERIAIS E MÉTODOS .....	41
5.2.1	Material Vegetal .....	41
5.2.2	Análise das Concentrações de Cafestol e Caveol por CLAE .....	42

5.2.3	Identificação e Validação de Cyt P450 de Coffea e Seleção de Genes Candidatos .....	44
5.2.4	Desenho de Oligonucleotídeos para qPCR .....	44
5.2.5	Extração de RNA .....	44
5.2.6	Análise Transcricional por qPCR .....	44
5.3	RESULTADOS E DISCUSSÃO .....	46
5.3.1	Quantificação de Cafestol e Caveol por CLAE .....	46
5.3.2	Seleção e Análise dos Genes Candidatos para a Biossíntese de Caf e Cav .....	49
5.3.3	Perfil Transcricional dos Genes Candidatos de Cyt P450 .....	51
5.4	REFERENCIAL BIBLIOGRÁFICO .....	56
<b>6</b>	<b>CONCLUSÃO E PERSPECTIVAS</b> .....	<b>62</b>
	<b>REFERÊNCIAS BIBLIOGRÁFICAS</b> .....	<b>64</b>
	<b>ANEXOS</b> .....	<b>68</b>
	ANEXO A - Contigs formados para Cyt P450 e suas sequências.....	69
	ANEXO B - Contigs formados para Cyt P450 e suas sequências.....	70
	ANEXO C - Árvore Filogenética com as Cyt P450 de Coffea spp. e A. thaliana.....	71
	ANEXO D - Artigo Aceito pela revista Coffee Science em Janeiro de 2012.....	72
	ANEXO E - Cromatograma da quantificação de Cafestol .....	87
	ANEXO F - Gráfico de porcentagem de similaridade entre as sequências nucleotídicas dos 7 genes candidatos.....	88

## 1 INTRODUÇÃO

O Brasil é o maior produtor e exportador mundial de café, além de possuir o segundo maior mercado consumidor. O aumento do consumo da bebida nos últimos anos tem sido associado, entre outros fatores, a maior qualidade da bebida aliada a bons preços de mercado, assim como às suas propriedades funcionais e nutraceuticas (ABIC, 2011; CONAB, 2011).

Dois importantes compostos nutraceuticos da bebida do café são os diterpenos cafestol (Caf) e caveol (Cav). Diversos estudos relatam as relações positivas (propriedades antioxidantes e anticarcinogênicas) e negativas (aumento dos níveis de colesterol) destes diterpenos para a saúde humana, além de diversas outras funcionalidades, como a utilização pelas indústrias farmacêuticas e cosméticas (KURZOCK & SPEER, 2001; LEE & JEONG, 2007; HUBER et al., 2008). Eles são formados a partir da via metabólica de isoprenóides, cujos genes da etapa final da rota de biossíntese ainda são desconhecidos.

Entre os principais candidatos para a síntese, acúmulo e/ou degradação de Caf e Cav estão os genes de citocromo P450 (Cyt P450). Esta família gênica produz proteínas capazes de realizar processos de catálise, reações de oxido-redução e também possuem função de incorporar aminoácidos e/ou realizar modificações da estrutura das proteínas.

Inicialmente, foi realizada uma seleção e identificação *in silico* de sequências expressas (ESTs) para a superfamília gênica das Cyt P450 na Plataforma do Genoma Café brasileiro (<http://www.lge.ibi.unicamp.br/coffee>). Foram identificadas 84 P450s em *Coffea* spp., das quais foram selecionados oito genes candidatos, cujas expressões foram avaliadas por qPCR em folhas, flores e frutos em três estádios de desenvolvimento.

Foram também realizadas análises por cromatografia líquida de alta eficiência (CLAE) em plantas de *Coffea arabica* cv. IAPAR59 a fim de analisar o padrão de concentração desses diterpenos em botão floral, flor e no perisperma de frutos durante o seu desenvolvimento e maturação (30DAF à 210DAF – dias após a florada).

## 2 OBJETIVO GERAL

Identificar e caracterizar transcricionalmente genes de citocromo p450 (Cyt P450) potencialmente envolvidos na via de biossíntese de diterpenos em café e relacionar o perfil transcricional de Cyt P450 com o acúmulo/degradação de cafestol e caveol em diferentes tecidos de cafeeiro.

### 2.1 OBJETIVOS ESPECÍFICOS

- Identificar sequências nucleotídicas expressas de Cyt P450 em *Coffea* spp.
- Determinar a concentração de cafestol e caveol por CLAE em botões florais, flores e perisperma de frutos em 7 fases de desenvolvimento (30 DAF à 210DAF) e em frutos com 120DAF após aplicação de metil jasmonato.
- Analisar os perfis transcricionais de oito Cyt P450 por qPCR em amostras contrastantes para a concentração de diterpenos.

### 3 FUNDAMENTAÇÃO TEÓRICA

#### 3.1 A IMPORTÂNCIA SÓCIO-ECONÔMICA DA CAFEICULTURA

Atualmente o Brasil é o maior produtor e exportador mundial de café com uma safra aproximada de 40 milhões de sacas de 60 Kg no ano de 2010 (CONAB, 2011; ABIC, 2011). De acordo com o Conselho dos Exportadores de Café do Brasil (CECAFÉ, 2011), a exportação em 2010 chegou a 30.526 milhões de sacas de 60 Kg, dos quais 89% de *Coffea arabica* (Arábica), 9% de café solúvel e 2% de *Coffea canephora* (Robusta). Também possui o segundo maior mercado consumidor, atrás somente dos Estados Unidos (ABIC, 2011).

As áreas cafeeiras ocuparam em 2010 cerca de 2,3 milhões de hectares e estão concentradas no Centro-Sul do país, com destaque para quatro principais estados produtores: Minas Gerais, São Paulo, Espírito Santo e Paraná. A região Nordeste também apresenta plantações na Bahia, e na região Norte pode-se destacar o estado de Rondônia (CONAB, 2011). A produção de café Arábica se concentra em Minas Gerais, São Paulo, Paraná, Bahia e parte do Espírito Santo, enquanto o café Robusta é plantado principalmente no Espírito Santo e Rondônia (ABIC, 2011).

Além da importância econômica, o café possui uma grande importância social, pois cerca de 80% de sua produção é feita no sistema de agricultura familiar, gerando renda para milhares de empregados diretos e indiretos (VEZALLI, 2006).

#### 3.2 OS LIPÍDEOS DO CAFÉ

Nos últimos anos, a grande produção de café no mundo tem propiciado uma disponibilidade acentuada do produto no mercado. Para auxiliar a valorização e, conseqüentemente, a renda gerada na cadeia produtiva do café, campanhas têm sido realizadas junto aos produtores para agregação de valor ao produto e uma das estratégias é baseada no aumento da qualidade do café.

Entre os principais componentes químicos do grão de café estão os lipídeos, açúcares, sólidos solúveis, ácidos clorogênicos, cafeína e trigonelina (SPILLER, 1998; KY et al., 2001). Os lipídeos representam aproximadamente de 7 a

17% dos compostos químicos totais do café, em *C. arabica* os valores médios são de 15%, e para *C. canephora* 10% (KURZROK & SPEER, 2001). De acordo com FOLSTAR (1985), a fração lipídica é composta principalmente pelos triglicerídeos (75%) e pelos ésteres diterpênicos (10~20%), que são os principais constituintes do material insaponificável.

Os lipídeos desempenham um papel importante na qualidade da bebida do café, principalmente devido à hidrólise de triacilgliceróis e liberação dos ácidos graxos (unidade fundamental dos lipídeos) os quais são em seguida oxidados e inferem características “off-flavor” (sabor da bebida) (SPEER et al., 1993). Eles também têm um uso comercial importante, porque eles são largamente empregados para a aromatização de café solúvel e também são relatados na literatura como importantes para estabilizar a estrutura de membranas e proporcionar uma maior tolerância dos frutos ao frio quando encontrados em maior quantidade nestas estruturas (CLARKE, 1985).

A maior parte do óleo de café se concentra no endosperma do grão de café, enquanto somente uma pequena quantidade, a cera de café, encontra-se na camada exterior dos frutos (FOLSTAR, 1985).

### 3.3 ÁCIDOS GRAXOS DO CAFÉ

Os ácidos graxos representam 75% do óleo de café, estes estão principalmente presentes na forma esterificada em glicerol para formar triacilgliceróis (80% dos ácidos graxos). Dez a vinte por cento dos ácidos graxos são encontrados em diterpenos esterificados, uma pequena fração (2%) é encontrado em ésteres de esteróis e uma pequena porcentagem na forma livre (ácidos graxos livres) (SPEER & KÖLLING-SPEER, 2006).

Existe uma variação muito grande entre espécies de café em relação às proporções relativas dos quatro principais ácidos graxos em café verde (palmítico C16:0, esteárico C18:0, oléico C18:1 e linoléico C18:2) (DUSSERT et al., 2001). Os ácidos graxos encontrados na fração triacilglicerólica dos grãos de café verdes são principalmente C18:2 (30-50%), C16:0 (16-38%), C18:1 (7-18%) e C18:0 (4-10%). Através da comparação dos ácidos graxos em espécies de *Coffea* spp., é interessante notar que *C. arabica* contém mais C18:2 que *C. canephora*, 52-54% comparado com 43-49%, respectivamente (SPEER et al., 1993). Porém, *C.*

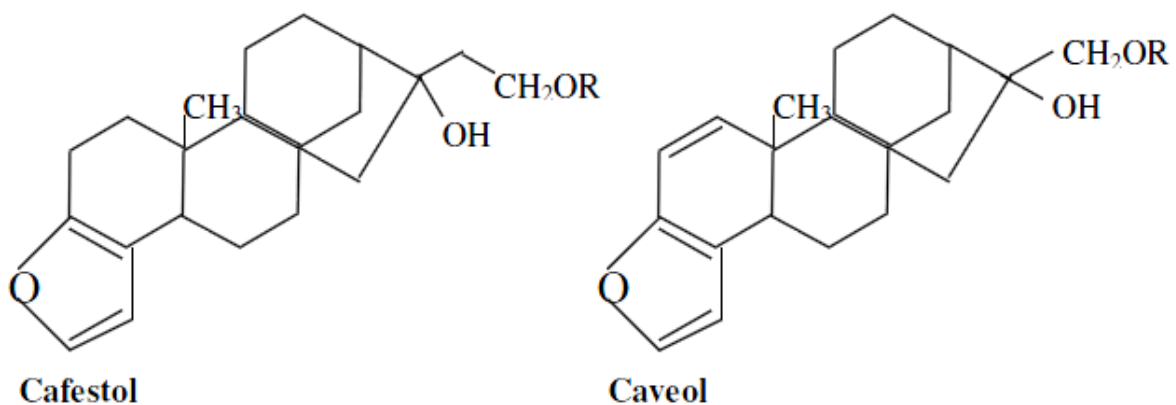
*canephora* contém mais C18:1 do que *C. arabica*, 9-14% e 6-8%, respectivamente. Autores sugerem que esta é uma característica que pode ser usada para dar uma primeira indicação da presença de *C. canephora* em misturas de café (SPEER et al., 1993; SPEER & KÖLLING-SPEER, 2006).

### 3.4 OS DITERPENOS DO CAFÉ

Os diterpenos (também conhecidos como ésteres de ácidos graxos) são produtos do metabolismo secundário das plantas. Estão relacionados ao mecanismo de defesa da planta, contra estresses bióticos e abióticos, e como atrativos para agentes polinizadores (McGARVEY & CROTEAU, 1995, THEIS & LEARDOU, 2003). Cafestol (Caf) e caveol (Cav) estão entre os diterpenos produzidos exclusivamente por plantas do gênero *Coffea*, e são de grande interesse em função dos seus efeitos fisiológicos (DIAS et al., 2010).

Suas moléculas possuem 20 átomos de Carbono (C<sub>20</sub>), que utilizam a via MVA (citosólica) e MEP (plastídica) para serem sintetizados. Os principais diterpenos em café são o Caf e o Cav, que são álcoois de diterpenos pentacíclicos (Figura 1) baseados na fusão de unidades de isopreno (C<sub>5</sub>) para formar o esqueleto caurano de vinte carbonos (SPEER & KOLLING-SPEER, 2006).

**Figura 1** - Fórmula estrutural do cafestol e caveol



Fonte: MASTEN & TICE et al., (1999).

O Café encontrado tanto em *C. arabica* (3-6 g/kg massa seca) quanto em *C. canephora* (1-2 g/kg massa seca), onde este diterpeno é

predominante. Porém, o Cav é específico de grãos de *C. arabica* (2 g/kg massa seca) (SPEER & KÖLLING-SPEER, 2006). Em grãos de café verdes, há somente uma pequena quantidade de diterpenos na forma livre, enquanto que a maior parte é encontrada na forma esterificada com ácidos graxos (radical), principalmente conjugados com ácido palmítico (C16:0) e ácido linoléico (C18:2) (SPEER et al., 1993). Esta situação contrasta com o restante da composição de ácidos graxos presentes em grãos verdes onde o ácido linoléico é o principal ácido graxo encontrado (SPEER et al., 1993; DUSSERT et al., 2001).

Análises de diterpenos nos tecidos de cafeeiros, demonstraram que os maiores picos de concentrações de Cav (*C. arabica* cv IAPAR-59) foram detectadas no perisperma (516 mg/100 g) aos 83DAF e no endosperma (589 mg/100 g) aos 240DAF, enquanto que em folhas e em café torrado os níveis desses diterpenos foram menores (DIAS et al., 2010). Para cafestol, o maior pico de maior concentração (304 mg/100 g) foi detectada no endosperma aos 83 DAF (Tabela 1).

**Tabela 1** - Picos de quantificação dos diterpenos nos tecidos de frutos e folha<sup>a</sup>

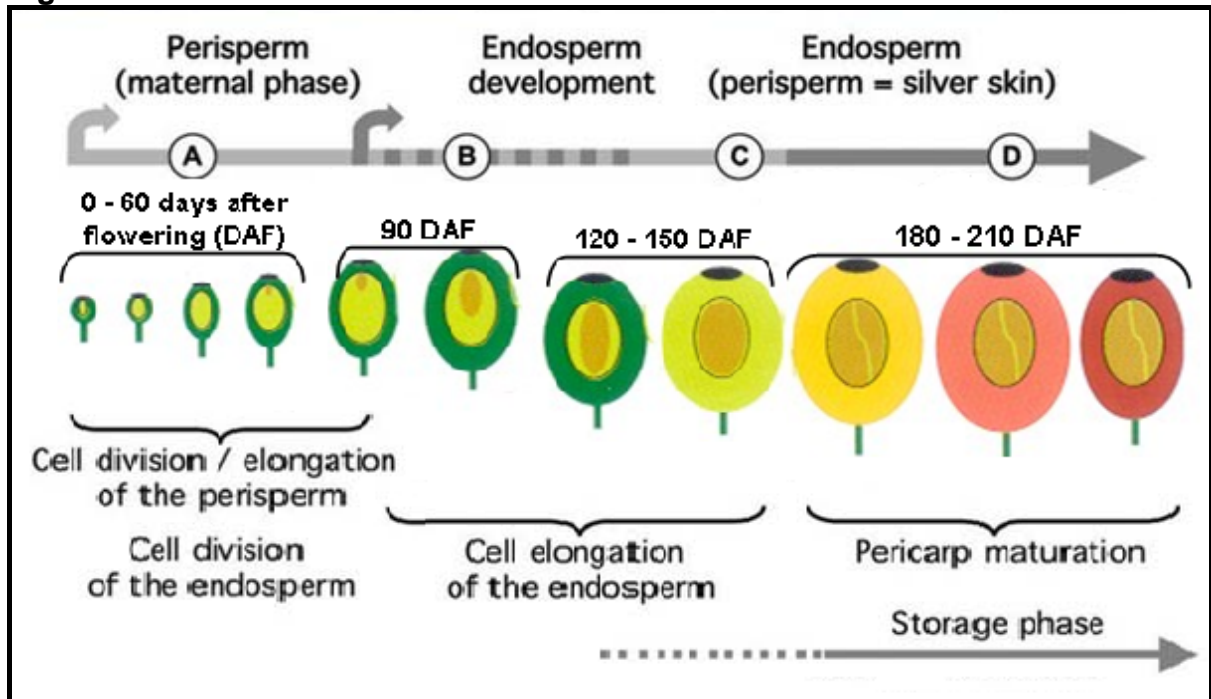
Diterpenos	Amostra	Perisperma	Endosperma	Polpa <sup>b</sup>	Folha <sup>c</sup>
Caveol	<i>C. arabica</i>	516 ± 68	589 ± 31	Nd	Nd
	<i>C. canephora</i>	Nd	Nd	Nd	Nd
Cafestol	<i>C. arabica</i>	132 ± 4	304 ± 11	49 ± 2	45 ± 2
	<i>C. canephora</i>	109 ± 4	94 ± 3	36 ± 2	Nd

<sup>a</sup> = Média (mg/100 g de amostra) para 5 repetições ± desvio-padrão (coeficiente de variação, %). Nd = não detectado. <sup>b, c</sup> Valores estimados pela equação da curva pois estão abaixo do menor ponto. Modificado de DIAS et al. (2010).

Ainda segundo DIAS et al. (2010), a produção de Caf e Cav ocorre principalmente, no perisperma e ao longo do desenvolvimento do fruto esses diterpenos são transferidos para o endosperma, assim como o que ocorre para o metabolismo da sacarose (GEROMEL et al., 2006). Segundo SRIDEVI et al., (2010) provavelmente os níveis de Caf e Cav aumentam ao longo do desenvolvimento do fruto, levando a um maior acúmulo nos frutos maduros (endosperma - 60% dos diterpenos totais) de *C. arabica*. O fruto de café possui um perisperma abundante durante as primeiras fases do seu desenvolvimento dando lugar ao endosperma durante a formação do grão de café (figura 2). No fruto maduro o perisperma se

torna residual (película prateada) e o endosperma é o tecido principal do grão (DE CASTRO & MARRACCINI, 2006).

**Figura 2** - Estádios de desenvolvimento do fruto de cafeeiro



Fonte: DE CASTRO & MARRACCINI et al., 2006.

Análises fenotípicas da composição de diterpenos entre diferentes espécies de *Coffea arabica* (com vários genótipos para cada espécie) mostraram que existe uma variabilidade significativa intraespecífica de Caf e Cav nos cafeeiros de *C. arabica* (POT et al., 2008; KITZBERGER et al., 2010). Estes estudos indicam que além da variabilidade ambiental, pode existir um controle genético relacionado com a síntese desses compostos.

Além das diferenças entre os teores de diterpenos detectados entre as espécies e dentro das espécies, estas quantidades podem ser diferentes de acordo com o modo de preparo da bebida (Tabela 2).

No tipo expresso os teores de diterpenos podem ser de 5 a 15 vezes maior que em bebida filtrada, pois o filtro de papel tende a retê-los (SPEER & KOLLING-SPEER, 2006; NAIDOO et al., 2011).

**Tabela 2** - Teores de cafestol e caveol de acordo com a forma de preparo<sup>a</sup>

<b>TIPO</b>	<b>Cafestol</b>	<b>Caveol</b>
<b>Turco/Grego</b>	3,12 ± 2,56	3,9 ± 3,9
<b>French Press</b>	2,8 ± 0,96	4,4 ± 2,1
<b>Escandinavo (evaporado)</b>	2,4 ± 2,24	3,9 ± 3,4
<b>Italiano (espresso)</b>	1,2 ± 0,8	1,8 ± 1,3
<b>Filtrado (papel)</b>	0,08 ± 0,0	Nd

<sup>a</sup> = Média ± desvio padrão, em mg de diterpenos por xícara (120mL); nd= dados não disponíveis. Modificado de URGERT et al., (1995) e NAIDOO et al., 2011.

Os diterpenos têm importância comprovada em diversos trabalhos que relatam as funções biológicas benéficas dos diterpenos do café, dentre eles a proteção contra alguns insetos e a atração de polinizadores (THEIS & LEARDAU, 2003). A indústria farmacêutica utiliza o óleo do café como filtro solar e a indústria de alimentos associou a valorização da qualidade da bebida (aroma) do café aos lipídeos (GROLLIER & PLESSIS, 1998). Com relação à saúde humana, alguns estudos observaram que estes diterpenos podem induzir a degradação de substâncias tóxicas proporcionando ação protetora contra a aflatoxina B1 e acroleína (CAVIN et al., 1998, HIGGINS et al., 2008), outros estudos atribuíram propriedades anticarcinogênicas (CAVIN et al., 2002; HUBER et al., 2008) e antioxidantes (LEE & JEONG, 2007) aos mesmos. Por outro lado algumas pesquisas inferiram que o Caf pode causar o aumento do nível de colesterol em humanos através da ingestão da bebida (URGERT et al., 1995; POSTE et al., 2000; NAIDOO et al., 2011).

### 3.5 BIOSÍNTESE DE DITERPENOS

#### 3.5.1 Síntese de Isoprenóides

As plantas sintetizam uma enorme variedade de metabólitos que podem ser classificados em dois grupos: metabólitos primários, com participação em processos essenciais no metabolismo da planta, e metabólitos secundários com

participação ativa em importantes funções ecológicas da planta (CROTEAU et al., 2000).

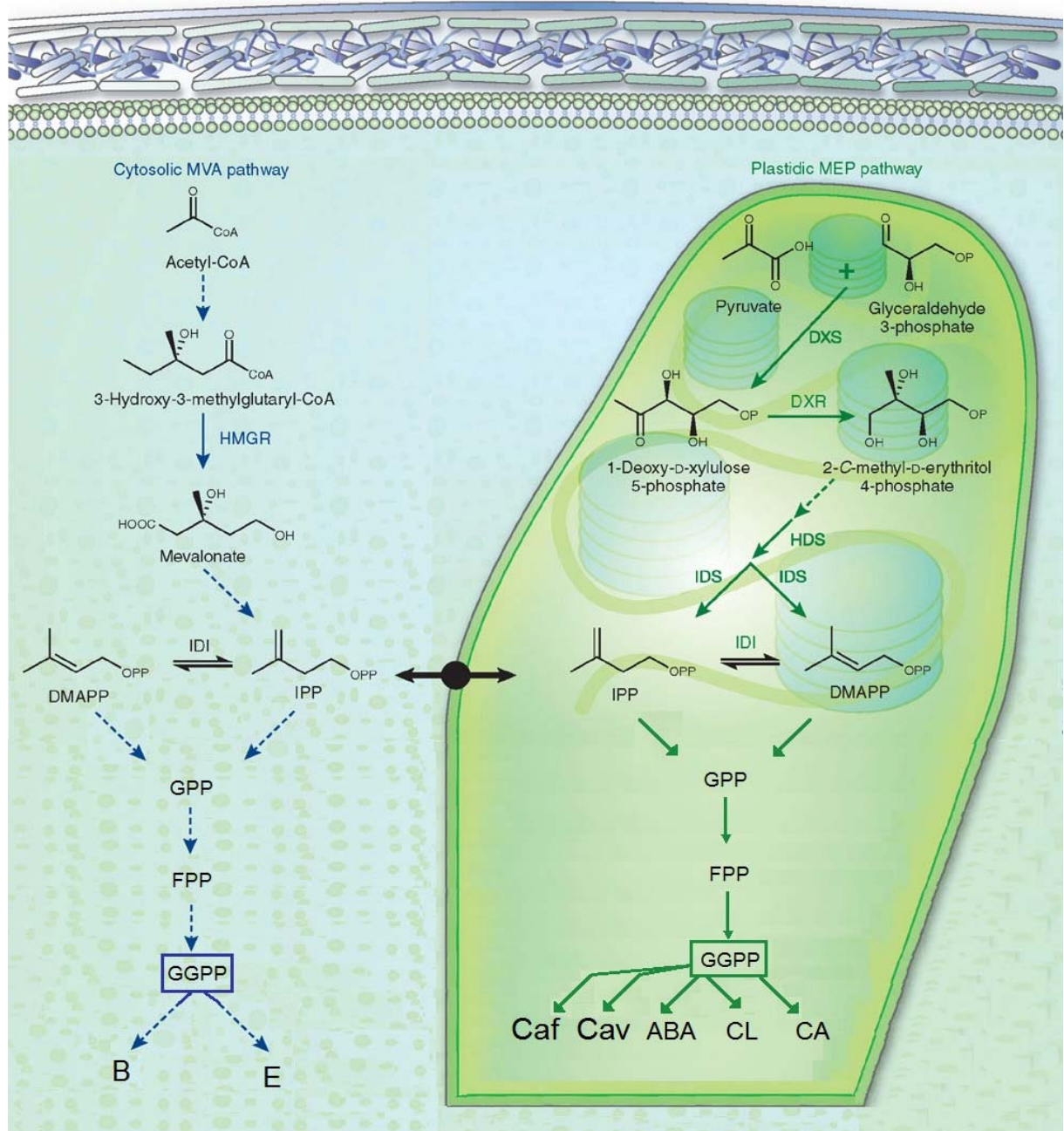
Os isoprenóides, também chamados de terpenóides, são sintetizados em todos os organismos principalmente em plantas (aproximadamente 10 mil diferentes componentes relatados) (CROTEAU et al., 2000). Muitos isoprenóides presentes em plantas atuam como metabólitos primários, com papel na respiração, fotossíntese, crescimento e desenvolvimento. Contudo, a maioria dos isoprenóides são metabólitos secundários, que atuam tanto na proteção de plantas contra herbívoros e patógenos microbianos, como na atração de agentes polinizadores (CROTEAU et al., 2000).

Segundo WILLE et al. (2004), as vias bioquímicas de isoprenóides (Figura 3) produzem diversos compostos químicos com funções diferentes, por exemplo, componentes de membranas (esteróis), pigmentos fotossintéticos (carotenóides e clorofilas) e hormônios (giberelinas). Os terpenóides são sintetizados através da condensação de unidades de isopreno, que são constituídos por moléculas de isopentenil difosfato (IPP) e o isômero dimetilalil difosfato (DMAPP), com 5 átomos de carbonos (C<sub>5</sub>).

Estudos indicam que a via MVA atua como precursora de farnesil difosfato (FPP, C<sub>15</sub>), sesquiterpenos e triterpenos e a via MEP atua como precursora de geranyl difosfato (GPP, C<sub>10</sub>), geranylgeranyl difosfato (GGPP, C<sub>20</sub>), monoterpenos, diterpenos (Caf e Cav) e tetraterpenos (LANGE & GHASSEMIAN, 2003).

A presença exclusiva de cafestol e caveol no gênero *Coffea* sugere a participação de enzimas específicas capazes de reconhecer e modificar o grupo caurano levando à produção de Caf e Cav, sendo que algumas destas enzimas ainda são desconhecidas até o presente momento.

**Figura 3 - Via MVA e MEP da biossíntese de isoprenóides e os seus respectivos compostos sintetizados**



HMGR= 3-hidroxi-metilglutaril-Coenzima A redutase; DXS= DXP sintase; DXR= DXP reductoisomerase; IDS= IPP/DMAPP sintase; IDI= IPP isomerase; IPP= isopentenil difosfato; FPP= farnesil difosfato; GPP= geranyl difosfato; GGPP= geranyl geranyl difosfato; Caf= cafestol; Cav= caveol; E= esteróis; B= Brassinoesteróis. ABA= ácido abscísico; CL= Clorofila; CA= Caroteno  
**Fonte:** (baseado em TUDYNSKI et al., 2001, WILLE et al., 2004 e ROBERTS, 2007).

Para o gênero *Coffea*, o estudo realizado por TISKI et al. (2011), descreveu e identificou o gene HMGR (3-hidroxi-metilglutaril-Coenzima A redutase) envolvido nas etapas iniciais da via MVA da formação de isoprenóides. Outro trabalho desenvolvido por FERREIRA et al. (2009), identificou e analisou os perfis transcricionais dos genes CPS (copalil fosfato sintase), KS (caureno sintase) e KO

(caureno oxidase) envolvidos nas etapas finais da biossíntese de diterpenos, sendo que este último pertence a família das Cyt P450 (WANG et al., 2012).

### 3 5.2 A Super Família das Cyt P450s

Os genes da superfamília das Cyt P450s estão presentes em diversos seres vivos, desde bactérias, protozoários, fungos, animais e plantas (NELSON, 2009). Atualmente, 5.100 seqüências de Cyt P450s foram anotadas em plantas, 1.461 em vertebrados, 2.137 em insetos, 2.960 em fungos, 1.042 em bactérias, 27 em *Archaea* e duas em vírus (NELSON, 2011). No total são 3651 Cyt p450s identificadas em 11 projetos genomas de plantas totalmente seqüenciados: *Arabidopsis thaliana*, mamão, videira, soja, tomate, arroz, *Brachypodium distachyon*, *Selaginella moellendorffii*, *Physcomitrella patens* (musgo), *Chlamydomonas reinhardtii* e *Volvox carteri* (alga verde). Além destes, existem 1449 seqüências de 255 espécies de plantas com genomas seqüenciados não completos (NELSON & WERCK-REICHHART, 2011), estas podem ser encontradas na página das P450s (<http://drnelson.uthsc.edu/CytochromeP450.html>; NELSON, 2009).

Elas são consideradas a maior família de enzimas no metabolismo das plantas sendo utilizadas para estudar a evolução de metabolismos de plantas e analisar processos de adaptação das mesmas (NELSON & WERCK-REICHHART, 2011). Os mesmos autores relatam que este grande número de enzimas são pontos chave para a produção de metabólitos e enzimas de sinalização e defesa da planta, além de participarem da polimerização de diversos componentes estruturais importantes (carotenóides, lignina, suberina, cutina, esteróis de membrana, hormônios, etc). Portanto, as Cyt P450s têm um papel fundamental para gerar a diversidade química encontrada nas plantas (NELSON & WERCK-REICHHART, 2011).

Para melhorar a compreensão da complexidade desta família, em *Arabidopsis thaliana* que possui 245 genes de Cyt P450 descritas, destas apenas 41 possuem função conhecida (SCHULER et al., 2006). Portanto, muitos estudos ainda são necessários para auxiliar a compreensão das funções das Cyt P450s nas diversas rotas metabólicas das plantas.

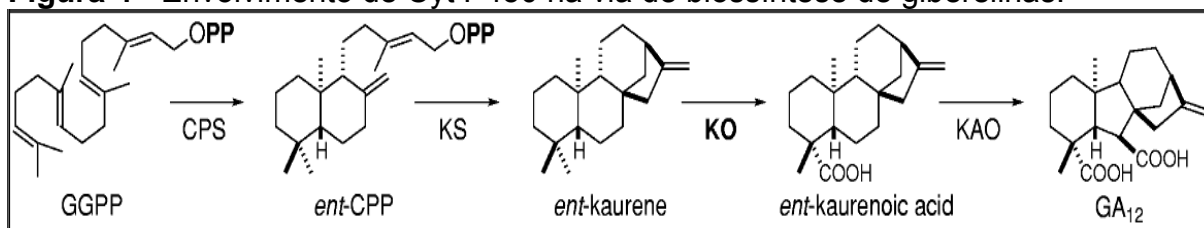
### 3.5.3 A Hipótese Para a Biossíntese de Cafestol e Caveol

As Cyt P450s codificam proteínas que possuem funções diferentes (monoxigenases, hidroxilases e oxidoredutases), as quais utilizam oxigênio e NAD(P)H para catalisar suas reações. A reação mais comum catalisada pelo citocromo P450 é uma reação monoxigenase, isto é, inserção de um átomo de oxigênio em um substrato orgânico (RH) com a utilização de elétrons, enquanto o outro átomo de oxigênio é reduzido à água como demonstra esta reação:  $RH + O_2 + 2H^+ + 2e^- \rightarrow ROH + H_2O$ .

Na literatura foram encontrados alguns trabalhos que relatam que nas vias de biossíntese de ácidos graxos há uma dependência de genes de Cyt P450 para catalisar a produção desses compostos (PINOT et al., 2007). Segundo os mesmos autores, alguns genes de Cyt P450 possuem função conhecida, como por exemplo, de hidroxilação de ácido láurico, mirístico e cáprico (CYP81B1) em *Helianthus tuberosus* e na hidroxilação do ácido palmítico em *Vicia faba* (CYP94A1).

Em *Arabidopsis thaliana* foi identificado um gene de Cyt P450 que faz hidroxilação na via metabólica de formação de carotenóides (DELLAAPPENNA et al., 2004). Outro estudo realizado por WANG et al. (2012) em *Oryza sativa*, demonstrou que a Cyt P450 (CYP701A8) é uma caureno oxidase (KO) e realiza hidroxilação na via de biossíntese de giberelinas (GA), convertendo o composto ent-caureno em ácido iso-caurenóico (KO) (Figura 4).

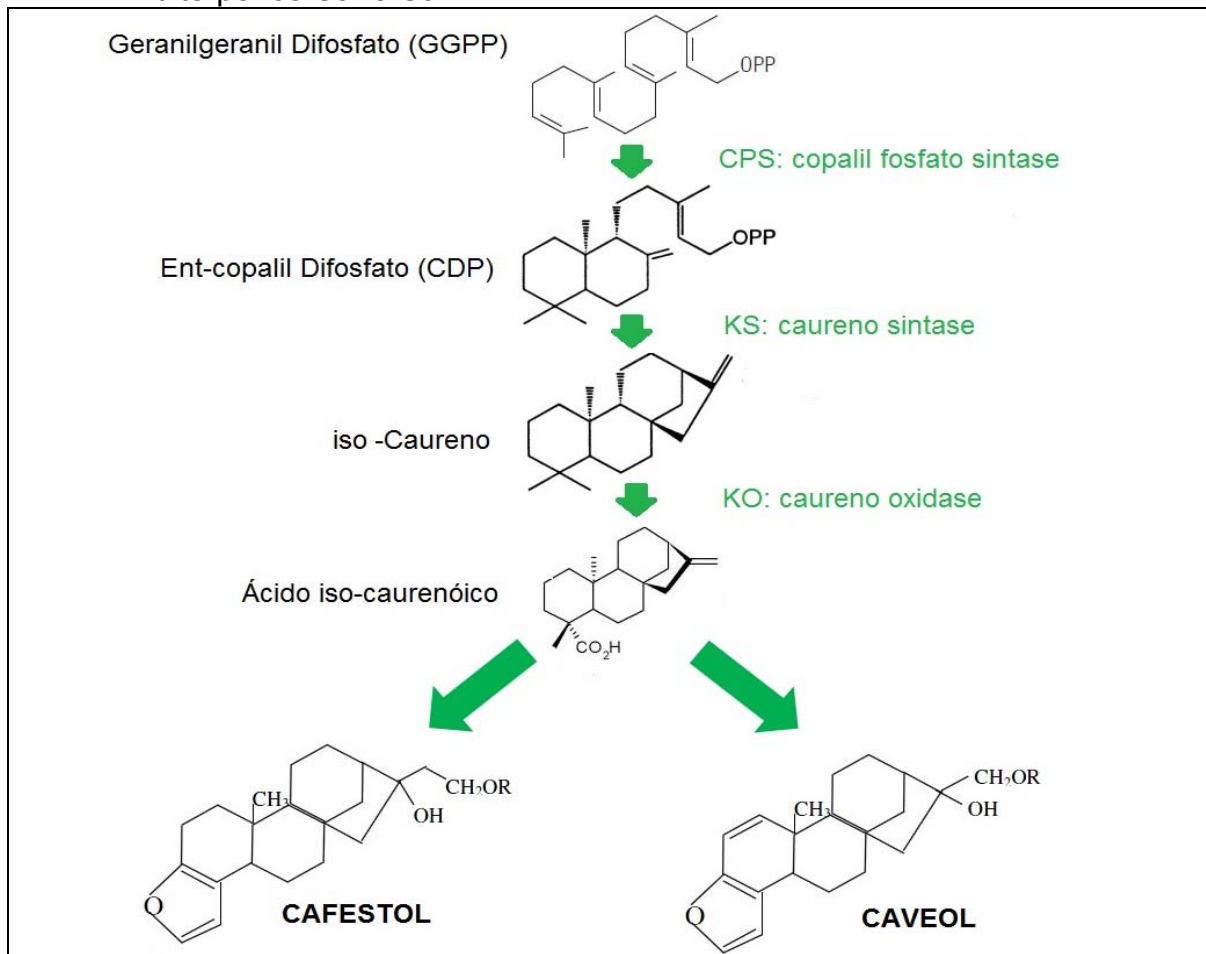
**Figura 4** - Envolvimento de Cyt P450 na via de biossíntese de giberelinas.



Fonte: (WANG et al., 2012).

Portanto de acordo com as funções descritas na literatura para as Cyt P450s (monoxigenases), é provável que a expressão destes genes possa interferir no acúmulo e/ou degradação de Caf e Cav, visto que as reações catalisadas por estes genes são compatíveis para a modificação estrutural da proteína ácido iso-caurenóico em Caf e/ou Cav (Figura 5).

**Figura 5** - Hipótese sobre a sequência das etapas finais da biossíntese dos diterpenos Caf e Cav.



**Fonte:** Baseado em WILLE et al., (2004), ROBERTS (2007) e WANG et al., (2012).

## 4 ARTIGO 1: A SUPER FAMÍLIA DAS CITOCROMOS P450s EM *COFFEA*

### 4.1 INTRODUÇÃO

Os genes da superfamília das Cyt P450s estão presentes em grandes grupos taxonômicos, desde bactérias, protozoários, fungos, plantas e em animais (NELSON, 2009). Atualmente, 5.100 seqüências de Cyt P450s foram anotadas em plantas, 1.461 em animais vertebrados, 2.137 em insetos, 2.960 em fungos, 1.042 em bactérias, 27 em *Archaea* e duas em vírus (NELSON, 2011). No total são 3.651 Cyt P450s identificadas em 11 projetos genomas de plantas totalmente seqüenciados: *Arabidopsis*, mamão, videira, soja, tomate, arroz, *Brachypodium distachyon*, *Selaginella moellendorffii*, *Physcomitrella patens* (musgo), *Chlamydomonas reinhardtii* e *Volvox carteri* (algas verdes). Além destes, existem 1449 seqüências de 255 espécies de plantas com genomas seqüenciados não completos (NELSON & WERCK-REICHHART, 2011), estas podem ser encontradas no site “Cytochrome P450 Homepage” (<http://drnelson.uthsc.edu/CytochromeP450.html>; NELSON, 2009).

Elas são consideradas a maior família de enzimas no metabolismo das plantas, tratando-se de pontos chave para a produção de metabólitos, hormônios, enzimas de sinalização e defesa da planta, além de participarem da polimerização de diversos componentes estruturais importantes (carotenóides, lignina, suberina, cutina, esteróis de membrana, etc), portanto, as Cyt P450s tem um papel fundamental para gerar a diversidade química que é a marca de plantas (NELSON & WERCK-REICHHART, 2011).

Esta superfamília pode ser dividida em duas famílias: tipo A (relacionados ao metabolismo secundário de plantas) e tipo não A (relacionados ao metabolismo primário); subdivididas em 11 clãs diferentes (10 para tipo A e 1 para tipo não A). Dentro dos clãs elas ainda podem ser agrupadas em subfamílias (NELSON, 2009; GUTTIKONDA, et al., 2010; NELSON & WERCK-REICHHART, 2011).

Apesar de sua importância, não há relatos na literatura de estudos de caracterização da família de Cyt P450 no gênero *Coffea*. Com base nos dados genômicos disponíveis até o momento, este trabalho identificou e caracterizou 84 citocromos P450s e as relacionou com enzimas já caracterizadas em plantas.

## 4.2 MATERIAIS E MÉTODOS

### 4.2.1 Identificação das Sequências de Cyt P450

A análise inicial para a busca das sequências de Cyt P450 de *Coffea* spp. foi feita através de uma busca por palavra chave, no banco de dados do Projeto Genoma Café (<http://www.lge.ibi.unicamp.br/coffee>) com o uso da palavra chave “P450” (ferramenta *blast* por *keyword*). Todos os cromatogramas das sequências resultantes foram alinhados através do software Codon Code Aligner (CCA) versão 3.5.7. As sequências nucleotídicas consensos dos contigs formados foram comparadas com seqüências do NCBI através da ferramenta *BlastX*, para verificar a presença das sequências em outras espécies de plantas.

A identificação de genes de Cyt P450 de cafeeiros foi feita baseada nos resultados de similaridade com sequências de espécies descritas anteriormente com *e-value* inferior à  $1e-20$  e presença de no mínimo 4 sequências resultantes da busca anotadas como sendo genes de Cyt P450.

Depois da confirmação de codificação das sequências proteicas putativas para Cyt P450, as sequências consensos foram comparadas na plataforma do Genoma Café por *BlastN*. Dessa forma, foi realizada uma busca mais acurada a fim de se obter todas as sequências relacionadas às Cyt P450 (processo de saturação). Foi selecionado um total de 2215 sequências e suas respectivas sequências foram utilizadas no alinhamento final das mesmas (todas estas possuíam o seu respectivo cromatograma).

A limpeza das sequências foi realizada por ferramentas do programa CCA, onde foram removidas as sequências dos vetores utilizados no seqüenciamento do Projeto Genoma Café (*trim vector*). Também foi realizada a remoção de cromatogramas de baixa qualidade (segundo parâmetros padrões dos programas phred e phrap, instalados dentro do CCA – *call base*). Foram eliminados os terminais de baixa qualidade das sequências (*clip ends*) e identificação de possíveis *heterozygous find indels* (picos inespecíficos, ou seja, o programa não foi capaz de identificar o nucleotídeo em determinadas localizações da seqüência por apresentarem mais de um pico) os quais também foram excluídos.

Após a limpeza das seqüências, estas foram alinhadas pelo Programa CCA e conseqüente formação de *contigs*. Para isto, alguns parâmetros

foram previamente definidos para o alinhamento e estes foram: percentual mínimo de identidade de 90% e sobreposição mínima de 20 pares de base.

#### 4.2.2 Validação e Caracterização das Cyt P450

Após a formação dos contigs, as seqüências consensos dos mesmos foram analisadas através da ferramenta *BlastX* no site do NCBI, para verificar se eles realmente se tratavam do gene de interesse (Cyt P450). Além disso, foi realizada uma busca para a presença de domínios específicos da super família das Cyt P450 (<http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml> - MARCHLER-BAUER et al., 2011) e a qual *cluster* do Projeto Genoma Café o contig correspondia.

Em seguida, as seqüências consensos dos contigs foram comparadas por *BlastX* ao genoma de *Arabidopsis thaliana* (TAIR - <http://www.arabidopsis.org>), a fim de fazer uma relação entre as P450s de *Coffea* e as já previamente anotadas nesta planta modelo.

#### 4.2.3 Árvore Filogenética

As sequências de cafeeiro e de *Arabidopsis thaliana* foram traduzidas eletronicamente em respectivas prováveis proteínas, alinhadas pelo programa clustalW que gera uma matriz de similaridade (THOMPSON et al., 1994) e esta foi utilizada para a obtenção da árvore filogenética no programa MEGA5 (TAMURA et al., 2011), pelo método de neighbor-joining, com base na matriz BLOSUM e distância p, com bootstrap de 1000 réplicas.

### 4.3 RESULTADOS E DISCUSSÃO

Na plataforma do Genoma Café brasileiro estão disponibilizadas um total de 267.533 ESTs de três espécies de cafeeiros, *Coffea arabica* (CA), *Coffea canephora* (CC) e *Coffea racemosa* (CR). Foram selecionadas 2.215 seqüências para os genes da superfamília das Cyt P450, 2.153 formaram 107 *contigs* no programa Codon Code Aligner (anexo 1 e 2), e 62 não alinharam com nenhuma seqüência (*singlets*).

Das sequências selecionadas, 1.246 são provenientes da biblioteca de CA, 930 da biblioteca de CC e 40 da biblioteca de CR, sendo que a maioria dos *contigs* apresentou um número inferior a 20 reads (anexos 1 e 2).

Os 107 *contigs* foram analisados (BlastX) contra o banco de dados do *GenBank* (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>), para verificar e validar a similaridade das Cyt P450s de café com as que estão caracterizadas em outras espécies (*Arabidopsis thaliana*, *Vitis vinifera*, *Nicotiana tabacum*, *Medicago truncatula*, *Lycopersicum lycopersicum* e *melongena*, *Glycine max*, dentre outras). A maioria dos *contigs* (84) apresentaram similaridade para Cyt P450 monoxigenases, enquanto que 7 *contigs* foram similares a Cyt P450 redutases (NADPH), 4 foram desconhecidos (*unknown protein*) e 12 apresentaram similaridade com outras proteínas (tabela 3).

**Tabela 3** - Classificação das Cyt P450 dos *contigs* de *Coffea* spp.

Resultado BlastX (NCBI)	Contigs de Coffea *
Cyt P450 Monoxigenases (84 contigs)	6, 7, 9, 10, 12, 14, 15, 16, 17, 18, 19, 20, 22, 23, 24, 25, 28, 29, 31, 32, 33, 35, 36, 37, 39, 40, 41, 42, 43, 45, 46, 47, 48, 49, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 63, 64, 65, 66, 67, 68, 69, 70, 72, 73, 74, 75, 76, 77, 80, 81, 82, 83, 84, 85, 86, 87, 89, 90, 91, 92, 93, 94, 95, 96, 97, 99, 100, 101, 102, 103, 104, 105 e 107
Cyt P450 (NADPH) redutases (7 contigs)	3, 4, 5, 13, 21, 38, 62
Proteínas com Função Desconhecidas (4 contigs)	11, 44, 71 e 78
Outras Proteínas (12 contigs)	1, 2, 8, 26, 27, 30, 34, 50, 79, 88, 98 e 106

\* Os números representam o seu respectivo *contig*.

A partir destes resultados de *BlastX* (NCBI) aliados a busca por *BlastX* no site de *Arabidopsis* (tabela 4) e procura por domínios conservados (NCBI), os *contigs* foram anotados com relação a quais famílias (tipo A ou não tipo A, relacionados ao metabolismo secundário e primário, respectivamente) e subfamílias das Cyt P450s elas pertenciam. Com isto, pode-se identificar 84 Cyt P450s de *Coffea* spp., sendo 33 da família tipo não A e 53 da família do tipo A (tabelas 5 e 6).

A maioria dos *contigs* encontrados apresentou domínio conservado com similaridade com a superfamília Cyt P450s (CL12078, pfam00067, COG2124 e PLN03112) e em alguns casos pode-se identificar domínios específicos para as subfamílias CYP90A (PLN02987), CYP94C (PLN02426) e domínios relacionados a

função de genes como flavonóide 3'-5'-hidroxilase (PLN00110), flavonóide 3'-monoxigenase (PLN02687), ácidos graxos (Omega-1)-hidroxilase/alcanos hidroxilases (PLN02169 e PLN03195), trans-cinamato-4-monoxigenase (PLN02394), epsilon hidroxilase (PLN02936), beta caroteno hidroxilase (PLN02738) e ent-kaurene oxidase (PLN02655).

**Tabela 4 - Classificação das Cyt P450 (BlastX - TAIR e NCBI/CDD)**

Contigs Café	TAIR	E-value	Genoma Café	E-value	Domínios Conservados (NCBI)		
	(AGI Locus)		(Cluster / ESTs)		CL	Pfam	PLN / CD / COG
Contig 1	AT5G59320 *	5E-21	Contig20472	0.0	não p450	não p450	não p450
Contig 2	ATMG00280 *	1.60E-02	Contig19629	0.0	não p450	não p450	não p450
Contig 3	ATG2G01021 *	0.11	Contig20717	4E-119	não CDD	não CDD	não CDD
Contig 4	ATG5G04810 *	1.8	Contig5600	4E-33	não CDD	não CDD	não CDD
Contig 5	AT4G35800 *	0.52	Contig4764	6E-19	não p450	não p450	não p450
Contig 6	AT1G12740	1E-143	Contig13999	0.0	CL12078	nd	nd
Contig 7	AT2G30490	0.0	Contig11795	0.0	CL12078	nd	nd
Contig 8	AT1G64720 *	1E-126	Contig12169	0.0	CD08870	no p450	no p450
Contig 9	AT3G26180	2E-52	Contig16441	0.0	nd	nd	nd
Contig 10	AT3G26300	3E-95	Contig13167	0.0	CL12078	nd	nd
Contig 11	AT4G01410 *	4E-59	Contig21234	0.0	CL12118	pfam03168 (4.74e-05)	nd
Contig 12	AT2G40890	0.0	Contig1106	0.0	nd	nd	nd
Contig 13	AT4G24520	0.0	Contig1545	0.0	nd	nd	CD06204 (1.16e-171)
Contig 14	AT5G07990	1E-102	Contig7504	0.0	CL12078	nd	PLN02687 (1.95e-56)
Contig 15	AT3G48310	1E-120	Contig2787	0.0	CL12078	nd	nd
Contig 16	AT3G14690	1E-140	Contig22798	0.0	CL12078	pfam00067 (7.68e-64)	nd
Contig 17	AT1G13080	4E-74	Contig4152	0.0	CL12078	nd	nd
Contig 18	AT5G04660	0.0	Contig6380	0.0	CL12078	nd	nd
Contig 19	AT2G45570	1E-119	Contig15765	0.0	CL12078	nd	nd
Contig 20	AT4G31500	1E-115	Contig16027	0.0	CL12078 (1.07e-119)	nd	nd
Contig 21	AT4G30210	0.0	Contig4343	0.0	CL00438 (2.07e-28)	pfam00667 (3.84e-79)	CD06204 (6e-161)
Contig 22	AT4G37360	1E-81	Contig21096	0.0	CL12078 (6.62e-65)	nd	nd
Contig 23	AT5G63450	3E-77	Contig19826	0.0	CL12078 (1.11E-125)	nd	nd
Contig 24	AT4G31940	2E-47	Contig7913	0.0	CL12078 (2.40e-32)	nd	nd
Contig 25	AT3G14690	1E-146	Contig927	0.0	CL12078 (1.15e-34)	pfam00667 (1.15e-34)	nd
Contig 26	AT1G78380 *	2E-14	Contig20472	0.0	não p450	não p450	não p450
Contig 27	AT4G14800 *	7E-11	Contig8673	9E-86	não p450	não p450	não p450
Contig 28	AT4G31940	1E-106	Contig9404	0.0	CL12078 (1.28e-94)	nd	nd
Contig 29	AT1G01280	0.0	Contig7019	0.0	CL12078 (0e+00)	nd	nd
Contig 30	AT1G73960 *	2.9	Contig17829	2E-94	não p450	não p450	não p450
Contig 31	AT5G10610	2E-80	Contig22859	0.0	CL12078 (8.45e-98)	nd	nd
Contig 32	AT5G36110	1E-152	Contig6875	0.0	CL12078 (6.68e-70)	nd	nd
Contig 33	AT3G48280	1E-141	Contig17636	0.0	CL12078 (4.26e-127)	nd	nd
Contig 34	AT2G23240 *	3E-15	Contig20472	0.0	não p450	não p450	não p450
Contig 35	AT4G37340	5E-46	Contig19760	0.0	CL12078 (1.11e-33)	nd	nd
Contig 36	AT1G64950	3E-77	Contig19822	0.0	CL12078 (1.71e-70)	nd	nd
Contig 37	AT5G42650	1E-136	Contig18902	0.0	CL12078 (3.85e-09)	pfam00067 (3.85e-09)	nd
Contig 38	AT4G30210	9E-66	Contig4343	0.0	CL06868 (5.44e-78)	nd	cd06204 (5.44e-78)
Contig 39	AT3G14690	1E-163	Contig15444	0.0	CL12078 (3.77e-77)	pfam00667 (3.77e-77)	nd
Contig 40	AT4G31970	1E-57	Contig7881	0.0	CL12078 (9.43e-56)	nd	nd
Contig 41	AT5G24910	1 e-106	Contig19459	0.0	CL12078 (8.83e-54)	pfam00667 (8.83e-54)	nd
Contig 42	AT1G12740	7E-53	Contig21160	0.0	CL12078 (7.13e-39)	nd	nd
Contig 43	AT3G26290	2E-56	Contig10746	0.0	CL12078 (3.92e-68)	nd	nd
Contig 44	AT5G50320 *	2.4	Contig20472	1E-160	não p450	não p450	não p450
Contig 45	AT3G14690	1E-139	Contig12077	0.0	CL12078 (4.29e-83)	pfam00667 (4.29e-83)	nd

\* = não é uma Cyt P450; \*\* = é uma Cyt P450 NADPH; ns = não significativo; nd = não detectado; não p450 = domínio não relacionado as Cyt P450s; não CDD: domínio conservado não presente na sequência.

**Tabela 4 - Classificação das Cyt P450 (BlastX - TAIR e NCBI/CDD) (continuação)**

Contig 46	AT5G07990	2E-85	Contig728	0.0	CL12078 (3.99e-128)		nd
Contig 47	AT5G24910	1E-32	Contig19459	0.0	CL12078 (1.44e-22)	pfam00667 (1.44e-22)	nd
Contig 48	AT1G31800	0.0	Contig12477	0.0	CL12078 (0e+0)	nd	nd
Contig 49	AT5G36110	2E-84	Contig14525	0.0	CL12078 (9.89e-53)	nd	nd
Contig 50	ns	ns	Contig1627	0.0	não p450	não p450	não p450
Contig 51	AT1G11600	9E-94	Contig6642	0.0	CL12078 (4.50e-43)	nd	nd
Contig 52	AT2G23190	2E-45	Contig5644	0.0	CL12078 (9.75e-33)	nd	nd
Contig 53	AT5G36110	1E-112	Contig15830	0.0	CL12078 (0e+0)	nd	nd
Contig 54	AT4G19230	0.0	Contig7244	0.0	CL12078 (1.12e-40)	nd	nd
Contig 55	AT3G48290	6E-44	Contig16946	0.0	CL12078 (1.12e-40)	nd	nd
Contig 56	AT5G05690	1E-158	Contig18601	0.0	CL12078 (1.12e-40)	nd	nd
Contig 57	AT3G53130	1E-75	Contig17135	0.0	CL12078 (1.34e-104)	nd	nd
Contig 58	AT4G39490	2E-69	Contig21261	0.0	CL12078 (1.42E-62)	nd	nd
Contig 59	AT3G26190	1E-24	Contig1735	0.0	CL12078 (4.61e-36)	nd	nd
Contig 60	AT3G52970	1E-38	Contig9419	0.0	CL12078 (2.01e-29)	nd	nd
Contig 61	AT3G03470	7E-30	Contig8152	0.0	CL12078 (2.83e-41)	nd	nd
Contig 62	AT4G24520	1E-80	Contig1545	0.0	CL06868 (4.55e-84)	nd	cd06204 (4.55e-84)
Contig 63	AT2G45510	4E-54	Contig16120	0.0	CL06868 (2.47e-45)	nd	nd
Contig 64	AT5G63450	4E-37	Contig19199	0.0	CL12078 (1.03e-26)	nd	PLN02426 (2.63e-28)
Contig 65	AT4G15110	2E-89	Contig8170	0.0	CL12078 (9.14e-64)	nd	PLN02738
Contig 66	AT3G25180	5E-58	Contig14870	0.0	CL12078 (1.76e-66)	nd	PLN02687 (1.76E-66)
Contig 67	AT3G14660	0.00004	Contig13627	0.0	CL12078 (1.85e-07)	nd	COG2124 (1.85e-07)
Contig 68	AT3G28740	2E-59	Contig12083	0.0	CL12078 (1.06e-47)	nd	PLN03112 (1.06e-47)
Contig 69	AT5G09443 *	0.16	Contig14524	0.0	CL12078 (5.41e-29)	nd	PLN03112 (5.41e-29)
Contig 70	AT1G73340	1E-76	Contig649	0.0	CL12078 (1.15e-35)	nd	PLN02987 (1.15e-35)
Contig 71	AT1G03130 *	0.32	Contig21004	0.0	não p450	não p450	não p450
Contig 72	AT1G75130	1E-101	Contig12667	0.0	CL12078 (5.60e-38)	pfam00667 (1.12e-55)	COG2124 (5.60e-38)
Contig 73	AT1G13110	0.004	Contig2401	0.0	CL12078 (2.57e-56)	nd	PLN03112 (2.57e-56)
Contig 74	AT2G45970	1E-168	Contig11225	0.0	CL12078 (1.16e-113)	nd	PLN03195 (1.16e-113)
Contig 75	AT3G48310	0.014	Contig12382	0.0	CL12078 (6.33e-68)	nd	PLN02687 (6.33e-68)
Contig 76	AT3G48290	7E-32	Contig9403	0.0	CL12078 (1.09e-40)	nd	PLN03112 (1.09e-40)
Contig 77	AT1G74110	0.0000008	Contig8604	0.0	CL12078 (3.61e-42)	nd	PLN02687 (3.61e-42)
Contig 78	AT4G29080 *	2.2	Contig21668	7E-34	não p450	não p450	não p450
Contig 79	AT3G27330 *	0.51	Contig8787	0.0	não p450	não p450	não p450
Contig 80	AT2G40890	4E-48	Contig15280	0.0	CL12078 (1.31e-60)	nd	PLN02687 (1.31e-60)
Contig 81	AT4G37370	0.0006	Contig14368	0.0	CL12078 (7.97e-53)	nd	PLN02687 (7.97e-53)
Contig 82	AT3G26300	1E-39	Contig14107	0.0	CL12078 (1.11E-37)	nd	PLN02687 (1.11E-37)
Contig 83	AT4G11300 *	0.01	Contig15482	0.0	CL12078 (3.62e-44)	nd	PLN03112 (3.62e-44)
Contig 84	AT5G24910	1E-63	Contig2605	0.0	CL12078 (1.81e-25)	nd	COG2124 (1.81e-25)
Contig 85	AT5G10600	0.015	Contig9474	0.0	CL12078 ( 1.15e-80)	nd	PLN03112 ( 1.15e-80)
Contig 86	AT4G39480	4E-75	Contig5532	0.0	CL12078 ( 2.92e-84)	nd	PLN02169( 2.92e-84)
Contig 87	AT3G53130	6E-23	Contig17135	0.0	CL12078 ( 7.73e-157)	nd	PLN02936 ( 7.73e-157)
Contig 88	AT4G24520	0.000001	Contig1545	1E-105	nd	nd	nd
Contig 89	AT4G30360 *	0.69	Contig22452	0.0	CL12078 (2.83e-68)	nd	PLN02687 (2.83e-68)
Contig 90	AT2G45510	2E-75	Contig16375	0.0	CL12078 (5.90e-69)	nd	PLN03195 (5.90e-69)
Contig 91	AT4G00360	1E-30	Contig11225	0.0	CL12078 (4.42e-35)	nd	PLN03195 (4.42e-35)
Contig 92	AT2G45580	2E-45	Contig16029	0.0	CL12078 ( 1.06e-27)	nd	PLN03112 ( 1.06e-27)
Contig 93	AT4G15110	1E-20	A00-XX-AR1-031-H06-EB	0.0	CL12078 ( 1.87e-41)	nd	PLN02738 (1.87e-41)
Contig 94	AT4G12330	2E-47	Contig13895	0.0	CL12078 (6.45e-50)	nd	PLN00110 (6.45e-50)
Contig 95	AT1G50550	0.01	Contig10425	0.0	CL12078 (2.34e-42)	nd	PLN02687 (2.34e-42)
Contig 96	AT3G26330	1E-35	Contig21940	0.0	CL12078 (1.78e-39)	nd	PLN02687 (1.78e-39)
Contig 97	AT1G77580	0.59	Contig5534	0.0	CL12078 (1.32e-40)	nd	PLN02687 (1.32e-40)
Contig 98	AT5G47370 *	2.1	Contig19657	2E-56	não p450	não p450	não p450
Contig 99	AT5G27925 *	0.14	Contig9952	0.0	CL12078 ( 5.49e-42)	nd	PLN03112 ( 5.49e-42)
Contig 100	AT4G36220	2E-36	Contig22515	0.0	CL12078 (4.07e-43)	nd	PLN02687 (4.07e-43)
Contig 101	AT3G48300	0.0000001	Contig12587	0.0	CL12078 (6.55e-48)	nd	PLN02687 (6.55e-48)
Contig 102	AT3G01900	9E-15	C00-XX-SE4-093-C01-TL.	0.0	CL12078 (1.63e-17)	nd	PLN03195 (1.63e-17)
Contig 103	AT4G15380	0.026	Contig20420	0.0	CL12078 ( 6.02e-26)	nd	PLN03112 ( 6.02e-26)
Contig 104	AT3G26290	4E-13	Contig2915	0.0	não p450	não p450	não p450
Contig 105	AT5G45220 *	0.28	Contig5795	0.0	CL12078 ( 5.49e-11)	nd	PLN03195 ( 5.49e-11)
Contig 106	AT3G17340 *	6.9	ns	ns	CL12078 (6.55e-48)	nd	PLN02687 (6.55e-48)
Contig 107	AT5G25900	0.013	nd	nd	CL12078 (0.0)	pfam00667 (4.53e-64)	PLN02655 (0.0)

\* = não é uma Cyt P450; \*\* = é uma Cyt P450 NADPH; ns = não significativo; nd = não detectado; não p450 = domínio não relacionado as Cyt P450s; não CDD: domínio conservado não presente na sequência.

**Tabela 5 - Família Gênica das Cyt P450 de *Coffea* spp. do tipo não A**

Família: não tipo A		Café	Soja	Medicago	Arabidopsis	Arroz	Álamo	Uva	Musgo
Clã	Subfamília								
<b>CYP51</b>	CYP51	0	2	1	1	10	2	2	1
<b>CYP72</b>	CYP72	5	12	7	9	13	6	22	0
	<b>CYP709</b>	0	0	1	3	9	1	1	0
	CYP714	3	6	3	2	5	6	6	0
	CYP715	0	6	1	1	1	2	1	0
	CYP721	1	2	1	1	2	6	5	0
	CYP734	0	3	1	1	4	2	2	0
	CYP735	0	3	1	1	2	2	1	0
	CYP749	0	0	0	0	0	9	0	0
<b>CYP74</b>	CYP74	1	6	4	2	4	6	7	3
<b>CYP85</b>	CYP85	0	5	1	2	1	3	2	0
	CYP87	2	2	2	1	11	12	7	0
	CYP88	0	3	3	2	1	2	2	0
	CYP90	1	12	4	4	5	7	4	0
	<b>CYP702</b>	0	0	0	6	0	0	0	0
	CYP707	1	10	3	4	3	7	5	0
	<b>CYP708</b>	0	0	0	4	0	0	0	0
	CYP716	3	7	3	2	0	17	15	1
	CYP718	0	1	0	1	0	17	0	1
	CYP720	1	2	1	1	0	1	1	0
	CYP722	0	2	1	1	1	1	1	0
	CYP724	0	1	0	1	1	2	2	0
	<b>CYP728</b>	0	2	0	0	11	7	6	0
	CYP729	0	0	1	0	2	1	0	0
	<b>CYP733</b>	0	3	0	0	1	1	1	0
<b>Clã CYP86</b>	CYP86	2	9	3	11	5	8	6	2
	CYP96	1	7	5	13	12	9	5	0
	CYP704	4	5	14	3	7	6	6	6
<b>CYP97</b>	CYP97	6	5	4	3	3	3	3	3
<b>CYP710</b>	CYP710	0	2	1	4	4	1	1	2
<b>CYP711</b>	CYP711	0	4	2	1	5	2	1	0
<b>CYP727</b>	CYP727	0	1	0	0	1	2	1	0
<b>CYP746</b>	CYP746	0	0	0	0	0	0	0	1
<b>TOTAL</b>		<b>31</b>	<b>137</b>	<b>72</b>	<b>91</b>	<b>142</b>	<b>148</b>	<b>125</b>	<b>22</b>

Tabela baseada nos dados de GUTTIKONDA et al. (2010) com algumas modificações e o acréscimo das Cyt P450s encontradas por este trabalho (descritas na coluna "Café").

**Tabela 6 - Família Gênica das Cyt P450 de *Coffea* spp. do tipo A**

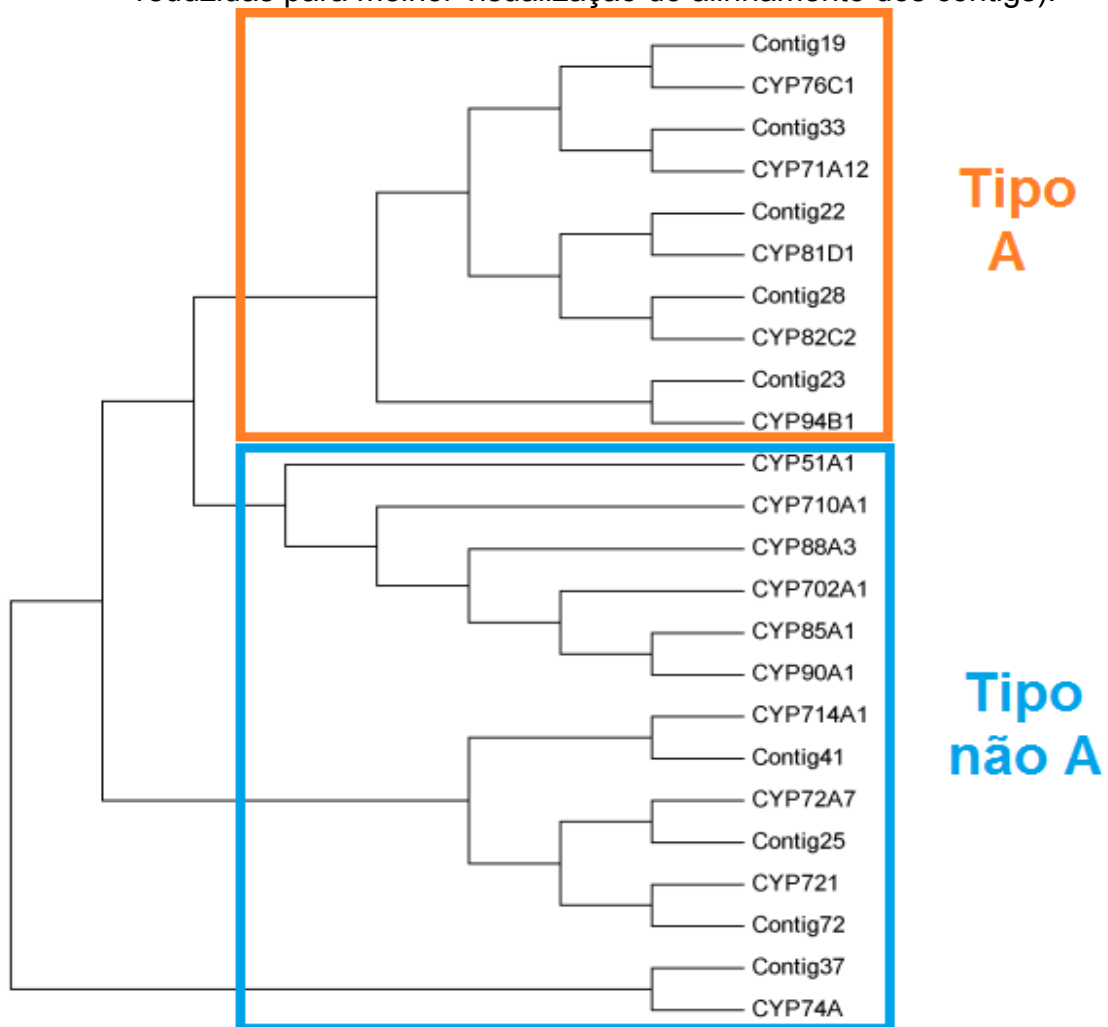
Família: Tipo A		Café	Soja	Medicago	Arabidopsis	Arroz	Álamo	Uva	Musgo
Clã	Subfamília								
<b>CYP71</b>	CYP71	17	55	37	52	84	25	24	0
	CYP73	1	3	1	1	3	3	3	4
	CYP75	3	7	0	1	3	3	11	0
	CYP76	4	14	6	8	29	13	24	0
	CYP77	2	4	2	5	2	3	2	0
	CYP78	1	11	1	6	8	10	7	3
	CYP79	0	5	3	7	4	4	9	0
	CYP80	0	0	0	0	0	6	6	0
	CYP81	8	12	5	18	12	28	21	0
	CYP82	7	24	10	5	0	10	34	0
	CYP83	0	12	9	1	0	5	0	0
	CYP84	2	3	3	2	3	3	3	0
	CYP89	2	8	9	7	14	10	14	0
	<b>CYP92</b>	<b>0</b>	<b>2</b>	<b>1</b>	<b>0</b>	<b>9</b>	<b>8</b>	<b>6</b>	<b>0</b>
	CYP93	0	13	8	1	3	4	4	0
	CYP94	2	14	4	6	18	13	9	2
	CYP98	1	2	1	3	2	5	1	1
	CYP99	0	0	0	0	2	0	0	0
	CYP701	1	2	1	1	5	1	1	1
	CYP703	1	1	1	1	1	1	1	3
	<b>CYP705</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>26</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>
	CYP706	1	3	1	7	4	5	9	0
	CYP712	0	2	1	2	0	9	2	0
CYP723	0	0	0	0	2	0	0	0	
<b>CYP736</b>	<b>0</b>	<b>12</b>	<b>1</b>	<b>0</b>	<b>0</b>	<b>6</b>	<b>8</b>	<b>0</b>	
<b>TOTAL</b>		<b>53</b>	<b>195</b>	<b>101</b>	<b>152</b>	<b>190</b>	<b>162</b>	<b>190</b>	<b>12</b>

Tabela baseada nos dados de GUTTIKONDA et al. (2010) com algumas modificações e o acréscimo das Cyt P450s encontradas por este trabalho (descritas na coluna "Café").

Os 107 contigs de *Coffea* spp. foram alinhados com sequências de Cyt P450 descritas na literatura para *Arabidopsis thaliana* (NELSON, 2009). A árvore filogenética resultante desta análise (anexo 3) corrobora os resultados encontrados por busca de similaridade (*BlastX*) no *GenBank* e reforçam a classificação dos contigs formados com suas respectivas subfamílias de Cyt P450 já descritas na literatura.

Para evidenciar a similaridade de alguns contigs de café com as Cyt P450s de *Arabidopsis thaliana*, foi gerada uma árvore filogenética parcial (Figura 6). Nesta figura observa-se que houve uma separação das duas famílias de Cyt P450 (metabolismo primário e secundário), sendo que o mesmo ocorreu para a análise completa de filogenia.

**Figura 6** - Árvore filogenética parcial das Cyt P450 (número de sequências reduzidas para melhor visualização do alinhamento dos contigs).



**Legenda:** superfamílias do tipo A (metabolismo secundário) em laranja e do tipo não A (metabolismo primário) em azul.

Com base nos dados de anotação, foi possível associar as Cyt P450 de cafeeiro com funções já descritas na literatura para algumas subfamílias (tabela 7). A maioria das Cyt P450 do tipo A encontradas em café pertencem à subfamília das CYP71 (17 contigs) relacionadas a reações de hidroxilação em estruturas carboxílicas como o limoneno 3- e 6-hidroxilases e o 2,4-dihydroxy-7-methoxy-1,4-benzoxazin-3-one (LUPIEN et al., 1999; FREY et al., 1997) e realizam sucessivas reações de hidroxilação/oxidação nas sínteses de tirosina-hidroxilase e de glucosídeos cianogênicos (BAK et al., 1998).

Foram encontrados quatro contigs envolvidos na biossíntese de fitohormônios: uma CYP90 (brassinosteróis – OHNISHI et al., 2006), duas CYP94 (jasmonoil L isoleucina – KOO et al., 2011) e uma CYP707 (ABA e GA – SAITO et al., 2004; KUSHIRO et al., 2004).

Dezessete contigs pertencem às subfamílias CYP72, CYP76, CYP74 e CYP81 (DIDIERJEAN et al., 2002; STUMPE & FEUSSNER, 2006; IMAISHI & MATUMOTO, 2007; PFALZ et al., 2009), as quais auxiliam as plantas na proteção contra herbívoros e na detoxificação de herbicidas das mesmas.

Entre as diversas subfamílias de Cyt P450 envolvidas principalmente com o metabolismo secundário foram identificados dois contigs para as CYP84 que fazem parte da via metabólica de ácido ferúlico 5-hidroxilase (KIM et al., 2006). Três contigs foram similares a CYP714 envolvidos na produção de alcalóides (ZHU et al., 2006), outros três contigs auxiliam na metabolização de ácidos graxos como CYP703 e CYP86 (IMAISHI et al., 1999; HÖFFER et al., 2008). Também foram encontrados três contigs similares às CYP75 as quais estão relacionados a síntese de flavonóides 3, 5-hidroxilase, na biossíntese de antocianina (SCHOENBOHM et al., 2000). Sete contigs foram identificados para CYP82 envolvidos na biossíntese de compostos voláteis em flores (CYP82 - LEE et al., 2010). Além desses foram encontrados dois contigs relacionados com as subfamílias CYP73 (cinamato 4-hidroxilase) e CYP98 (p-coumaroil-3-hidroxilase), ambas Cyt P450s envolvidas na via de biossíntese dos ácidos clorogênicos (MIZUTANI et al., 1997; CYP98 – KAI et al., 2006). Essas duas últimas subfamílias já possuem dados sobre os seus perfis transcricionais (LEPPELEY et al., 2003), diversidade nucleotídica (IVAMOTO et al., 2012) e estudos de mapas associativos com QTLs (lócos quantitativos) em café disponíveis na literatura (LEROY et al., 2011).

Existe ainda uma subfamília de CYP450 cujo gene produz uma proteína que exerce função relacionada com a biossíntese de diterpenóides, o gene CYP701 é o responsável pela produção da enzima ent-caureno oxidase (MORRONE et al., 2010). Segundo os mesmos autores, essa Cyt P450 catalisa oxidações sucessivas no composto iso-caureno para originar o ácido iso-caurenóico, este último provável precursor dos diterpenos Caf e Cav, responsáveis por diversas características nutracêuticas do café (1 contig identificado em *Coffea*).

Os dados de análise *in silico* deste trabalho também permitiram a identificação de oito genes candidatos à biossíntese de diterpenos que foram caracterizados no capítulo 2.

Todos esses dados abrem a possibilidade de estudos associativos e de mapeamento genético e físico desses genes (anexo 4) e podem auxiliar futuras investigações relacionadas a diversidade molecular das Cyt P450s de cafeeiros.

Tabela 7 - Prováveis funções das Cyt P450 de *Coffea*

Família	Subfamília	Contigs	Função	Bibliografia
Tipo A	CYP71	9, 10, 15, 17, 20, 33, 43, 55, 75, 76, 82, 83, 95, 96, 101, 103, 104	hydroxamic acid biosynthesis; cyanogenic glucoside biosynthesis, monoterpene biosynthesis; alkaloid biosynthesis; herbicides detoxification, ; camalexin biosynthesis, phytoalexin biosynthesis; mono and diterpene biosynthesis	Frey et al., 1997; Bak et al., 1998; Lupien et al., 1999; Schröder et al., 1999 ; Siminszky et al., 1999; Zhou et al., 1999; Ralston et al., 2001; Li et al., 2010a; Wang et al., 2001
	CYP73	7	cinnamate 4-hydroxylase	Mizutani et al., 1997; Ro et al., 2001
	CYP75	14, 80,100	flavonoid 3', 5'-hydroxylase	Schoenbohm et al., 2000
	CYP76	19, 60, 69, 92	herbicides detoxification and diterpenoid biosynthesis	Didierjean et al., 2002; Swaminathan, et al., 2009
	CYP77	18, 51	ormation of three mono-epoxides of alpha-linolenic acid and Chain hydroxylase for cutin synthesis for morphology of flower	Sauveplane et al., 2009 Li-Beisson et al., 2009
	CYP78	77	fatty acid omega-hydroxylase and control leaf growth	Imaishi et al., 2000; Anastasiou et al. ,2007
	CYP81	22, 31, 35, 52, 68, 81, 85, 97	biosynthesis of indolic glucosinolate	Pfalz et al., 2009; Kai et al., 2011
	CYP82	24, 28, 40, 66, 73, 89, 99	biosynthesis of floral homoterpene volatiles	Lee et al., 2010
	CYP84	46, 59	biosynthesis of ferulate-5-hydroxylase	Kim et al., 2006a
	CYP94	23, 64	jasmonate-responsive gene (JA-Ile-12-hydroxylase)	Koo et al., 2011
	CYP98	12	3'-hydroxylase for p-coumaryl shikimic/quinic acids (C3H)	Schoch et al., 2001; Kai et al., 2006
	CYP701	107	ent-kaurene oxidase	Morrone et al., 2010
	CYP703	29	Fatty acid omega-hydroxylase ; Sporopollenin synthesis; pollen synthesis,	Imaishi et al., 1999 Morant et al., 2007
	CYP706	1	phytoalexin biosynthesis	Luo et al., 2001
Tipo não A	CYP72	16, 25, 39, 45, 67	BL 26-hydroxylase; omega-hydroxylation of the herbicide pelargonic acid	Neff et al., 1999; Imaishi & Matumoto, 2007
	CYP714	41, 47 ,84	alkaloid biosynthesis	Zhu et al., 2006
	CYP74	37	hydroperoxide lyase (HPL) and allene oxide synthase (AOS)	Bate et al., 1998; Stump & Feussner, 2006
	CYP90	56	BR 22-hydroxylase and BR 23-hydroxylase	Ohnishi et al., 2006
	CYP707	54	8'-hydroxylase for ABA inactivation, Enhancement of ABA catabolism	Saito et al. 2004; Kushiro et al. 2004
	CYP716	32, 49, 53	Multifunctional Oxidases in Triterpenoid Biosynthesis formation of protopanaxadiol during ginsenoside biosynthesis	Fukushima et al., 2011 Han et al., 2011
	CYP720	70	resin acid biosynthesis	Hamberger et al., 2011.
	CYP86	74, 91	fatty acid omega-hydroxylase	Höffer et al., 2008
	CYP96	86	Formation of secondary alcohols and ketones in cuticular wax of stem, acyl CoA reductase	Greer et al., 2007
	CYP704	63, 90, 102, 105	Fatty acid and sporopollenin biosynthesis (pollen)	Li et al., 2010b
	CYP97	48, 57, 58, 65, 87, 93	b -ring and ε -ring hydroxylase on carotenes	Kim and DellaPenna, 2006; Tian et al., 2004.

#### 4.4 REFERENCIAL BIBLIOGRÁFICO

ANASTASIOU, E.; KENZ, S.; GERSTUNG, M.; DANIEL MACLEAN, D.; TIMMER, J.; FLECK, C.; LENHARD, M. Control of plant organ size by KLUH/CYP78A5-dependent intercellular signaling. **Developmental Cell**, v.13, p.843-856, 2007.

BAK, S.; OLSEN, C. E.; HALKIER, B. A.; MOLLER, B. L.; Transgenic *Tobacco* and *Arabidopsis* plants expressing the two multifunctional *Sorghum* cytochrome P450 enzymes, CYP79A1 and CYP71E1, are cyanogenic and accumulate metabolites derived from intermediates in dhurrin biosynthesis. **Plant Physiology**, v.123, p.1437-1448, 2000.

BATE, N. J.; SIVASANKAR, S.; MOXON, C.; RILEY, J. M. C.; THOMPSON, J. E.; ROTHSTEIN, S. J. Molecular characterization of an *Arabidopsis* gene encoding hydroperoxide lyase, a cytochrome P-450 that is wound inducible. **Plant Physiology**, v.117, p.1393-1400, 1998.

DIDIERJEAN, L.; GONDET, L.; PERKINS, R.; LAU, S. C.; SCHALLER, H.; O'KEEFE, D. P.; WERCK-REICHHART, D. Engineering herbicide metabolism in *Tobacco* and *Arabidopsis* with CYP76B1, a cytochrome P450 enzyme from Jerusalem artichoke. **Plant Physiology**, v.130, p.179-189, 2002.

FREY, M.; CHOMET, P.; GLAWISCHNIH, E.; STETTNER, C.; GRÜN, S.; WINKLMAIR, A.; WISENREICH, W.; BACHER, A.; MEELEY, R. B.; BRIGGS, S. P.; SIMCOX, K.; GIERL, A. Analysis of a chemical plant defense mechanism in grasses. **Science**, n.277, p.696-699, 1997.

FUKUSHIMA, E. R.; SEKI, H.; OHYAMA, K.; ONO, E.; UMEMOTO, N.; MIZUTANI, M.; SAITO, K.; MURANAKA, T. CYP716A subfamily members are multifunctional oxidases in triterpenoid biosynthesis. **Plant and Cell Physiology**, v.52, n.12, p.2050-2061, 2011.

GREER, S.; WEN, M.; BIRD, D.; WU, X.; SAMUELS, L.; KUNST, L.; JETTER, R.; The cytochrome P450 enzyme CYP96A15 is the midchain alkane hydroxylase responsible for formation of secondary alcohols and ketones in stem cuticular wax of *Arabidopsis*. **Plant Physiology**, v.145, p.653-667, 2007.

GUTTIKONDA, S. K.; TRUPTI, J.; BISHT, N. C.; CHEN, H.; AN, Y. C.; PANDEY, S.; XU, D.; YU, O.; Whole genome co-expression analysis of soybean cytochrome P450 genes identifies nodulation-specific P450 monooxygenases. **BMC Plant Biology**, v. 10, n.243, 2010.

HAMBERGER, B.; OHNISHI, T.; HAMBERGER, B.; SÉGUIN, A.; BOHLMANN, J. Evolution of diterpene metabolism: sitka spruce CYP720B4 catalyzes multiple oxidations in resin acid biosynthesis of conifer defense against insects. **Plant Physiology**, v.157, p.1677-1695, 2011.

- HAN, J.; KIM, H.; KWON, Y.; CHOI, Y. The cytochrome P450 enzyme CYP716A47 catalyzes the formation of protopanaxadiol from dammarenediol-II during ginsenoside biosynthesis in *Panax ginseng*. **Plant and Cell Physiology**, v.52, n.12, p.2062-2073, 2011.
- HÖFFER, R.; BRIESEN, I.; BECK, M.; PINOT, F.; SCHEIBER, L.; FRANKE, R. The Arabidopsis cytochrome P450 CYP86A1 encodes a fatty acid  $\omega$ -hydroxylase involved in suberin monomer biosynthesis. **Journal of Experimental Botany**, v.59, n.9, p.2347-2360, 2008.
- IMAISHI, H.; MATSUMOTO, Y.; ISHITOBI, U.; OHKAWA, H. Encoding of cytochrome P450-dependent lauric acid monooxygenase by CYP703A1 specifically expressed in the floral buds of *Petunia hybrid*. **Bioscience, Biotechnology and Biochemistry**, v.63, n.12, p.2082-2090, 1999.
- IMAISHI, H.; MATSUO, S.; SWAI, E.; OHKAWA, H. CYP78A1 Preferentially expressed in development inflorescences of *Zea mays* encoded a cytochrome P450-Dependent lauric acid 12-monooxygenase. **Bioscience, Biotechnology and Biochemistry**, v.64, n.8, p.1696-1701, 2000.
- IMAISHI, H. MATUMOTO, S. Isolation and functional characterization in yeast of CYP72A18, a rice cytochrome P450 that catalyzes  $\omega$ -hydroxylation of the herbicide pelargonic acid. **Pesticide, Biochemistry & Physiology**, v.88, p.71-77, 2007.
- IVAMOTO, S. T.; POT, D.; LANNES, S. D.; DOMINGUES, D. S.; VIEIRA, L. G. E.; PEREIRA, L. F. P. Nucleotide diversity of genes involved in the chlorogenic acids biosynthesis. **Coffee Science**, 2012 – (aceito para publicação em Jan/2012).
- KAI, K.; SHIMIZU, B.; MIZUTANI, M.; WATANABE, K.; SAKATA, K. Accumulation of coumarins in *Arabidopsis thaliana*. **Phytochemistry**, v.67, n.4, p.379-386, 2006.
- KAI, K.; TAKAHASHI, H.; SAGA, H.; OGAWA, T.; KANAYA, S.; OHTA, D. Metabolomic characterization of the possible involvement of a cytochrome P450, CYP81F4, in the biosynthesis of indolic glucosinolate in Arabidopsis. **Plant Biotechnology**, v.28, p.379-385, 2011.
- KIM, J.; DELLAPENNA, D. Defining the primary route for lutein synthesis in plants: the role of *Arabidopsis* carotenoid  $\beta$ -ring hydroxylase CYP97A3. **Proceedings of the National Academy of Sciences of the United States of America**, v.103, n.9, p.3474-3479, 2006.
- KIM, J.; KIM, J.; BAEK, M.; PARK, Y.; WI, S. G.; CHO, J.; CHUNG, B. Y. Cloning, characterization, and expression of two cDNA clones for a rice ferulate-5-hydroxylase gene, a cytochrome P450-dependent monooxygenase. **Journal of Plant Biology**, v.49, n.3, p.200-204, 2006.
- KOO, A. J. K.; COOKE, T. F.; HOWE, G. A. Cytochrome P450 CYP94B3 mediates catabolism and inactivation of the plant hormone jasmonoyl-L-isoleucine. **Proceedings of the National Academy of Sciences of the United States of America**, v.108, n.22, p.9298-9303, 2011.

KUSHIRO, M.; OKAMOTO, M. NAKABAYASHI, K.; YAMAGISHI, K.; KITAMURA, S.; ASAMI, T.; HIRAI, N.; KOSHIBA, T.; KAMIYA, Y. NAMBARA, E. The *Arabidopsis* cytochrome P450 CYP707A encodes ABA 8'-hydroxylases: key enzyme in ABA catabolism. **The European Molecular Biology Organization Journal**, v.7, n.23, p.1647-1656, 2004.

LEE, S.; BADIEYAN, S.; BEVAN, D. R.; HERDE, M.; GATZ, C.; THOLL, D. Herbivore-induced and floral homoterpene volatiles are biosynthesized by a single P450 enzyme (CYP82G1) in *Arabidopsis*. **Proceedings of the National Academy of Sciences of the United States of America**, v.107, n.49, p.21205-21210, 2010.

LEPELLEY, M.; CHEMINADE, G.; TREMILLON, N.; SIMKIN, A.; CAILLET, V.; MCCARTHY, J. Chlorogenic acid synthesis in *Coffee*: an analysis of CGA content and real-time RT-PCR expression of HCT, HQT, C3H1, and CCoAOMT1 genes during grain development in *C. canephora*. **Plant Science**, v.172, n.5, p.978-996, 2007.

LEROY, T.; DE BELLIS, F.; LEGNATE, H.; KANAMURA, E.; GONZALES, G.; PEREIRA, L. F.; ANDRADE, A. C.; CHARMETANT, P.; MONTAGNON, C.; CUBRY, P.; MARRACINI, P.; POT, D.; KOCHKO, A. Improving the quality of African robustas: QTLs for yield-and quality-related traits in *Coffea canephora*. **Tree Genetics & Genomes**, v. 7, n. 4, p. 781-798, 2011.

LI, Z.; HAO, T.; YANG, Y. DENG, W.; Molecular cloning and expression analysis of a cytochrome P450 gene in tomato. **Plant Growth Regulation**, v.61, n.3, p.297-304, 2010a.

LI, H.; PINOT, F.; SAUVEPLANE, V.; WERCK-REICHHART, D.; DIEHL, P.; SCHREIBER, L.; FRANKE, R.; ZHANG, P.; CHEN, L.; GAO, Y.; LIANG, W.; ZHANG, D. Cytochrome P450 family member CYP704B2 catalyzes the hydroxylation of fatty acids and is required for anther cutin biosynthesis and pollen exine formation in rice. **The Plant Cell**, v.22, p173-190, 2010b.

LI-BEISSON, Y.; POLLARD, M.; SAUVEPLANE, V.; PINOT, F.; OHLOROGGE, J.; BEISSON, F. Nanoridges that characterize the surface morphology of flowers require the synthesis of cutin polyester. **Proceedings of the National Academy of Sciences of the United States of America**, v.96, n.4, p.1750-1755, 2009.

LUO, P.; WANG, Y.; WANG, G.; ESSENBERG, M.; CHEN, X. Molecular cloning and functional identification of cadinene-8-hydroxylase, a cytochrome P450 mono-oxygenase (CYP706B1) of cotton sesquiterpene biosynthesis. **Plant Journal**, v.28, p.95-104, 2001.

LUPIEN, S.; KARP, F.; WILDUNG, M.; CROTEAU, R. Regiospecific cytochrome P450 limonene hydroxylases from mint (*Mentha*) species: cDNA isolation, characterization, and functional expression of (-)-4S-limonene-3-hydroxylase and (-)-4S-limonene-6-hydroxylase. **Archives of Biochemistry and Biophysics**, v.368, n.1, p.181-192, 1999.

MARCHLER-BAUER, A. et al. CDD: A curated Entrez database of conserved domain alignments. **Nucleic Acids Research**, v.31, n.1, p.383-387, 2003.

MIZUTANI, M.; OHTA, D.; SATO, R. Isolation of a cDNA and a genomic clone encoding cinnamate 4-hydroxylase from *Arabidopsis* and its expression manner in planta. **Plant Physiology**, v.113, n.3, p.755-763, 1997.

MORANT, M.; JORGENSE, K.; SCHALLER, H.; PINOT, F.; MOLLER, B. L.; WERCK-REICHHART, S.; BAK, S. CYP703 is an ancient Cytochrome P450 in land plants catalyzing in-chain hydroxylation of lauric acid to provide building blocks for sporopollenin synthesis in pollen. **The Plant Cell**, v.19, p.1473-1487, 2007.

MORRONE, D.; CHEN, X.; COATES, R. M.; PETERS, R. J. Characterization of the kaurene oxidase CYP701A3, a multifunctional cytochrome P450 from gibberellin biosynthesis. **Biochemical Journal**, v.431, p.337-344, 2010.

NEFF, M. M.; NGUYEN, S. M.; MALANCHARUVIL, E. J.; FUJIOKA, S.; NOGUCHI, T.; SETO, H.; TSUBUKI, M.; HONDA, T.; TAKATSUTO, S.; YOSHIDA, S.; CHORY, J. BAS1: A gene regulating brassinosteroid levels and light responsiveness in *Arabidopsis*. **Proceedings of the National Academy of Sciences of the United States of America**, v.96, p.15316-15323, 1999.

NELSON, D. R.; The Cytochrome P450 Homepage. Human Genomics, v.4, n.1, p.59-65, 2009. Disponível em: <<http://drnelson.uthsc.edu/cytochromeP450.html>>. Acesso em: 21 ago 2011.

NELSON, D.R. Progress in tracing the evolutionary paths of cytochrome P450. **Biochimica et Biophysica (BBA) – Proteins & Proteomics**, v.1814, n.1, p.14-18, 2011.

NELSON, D. R.; WERCK-REICHHART, D. The Cytochrome P450-centric view of plant evolution. **The Plant Journal**, v.66, p.194-211, 2011.

OHNISHI, T.; SZATMARI, A.; WATANABE, B.; FUGITA, S.; BANCOS, S.; KONGZ, C.; LAFOS, M.; SHIBATA, K.; YOKOTA, T.; SAKATA, K.; SZEKERES, M.; MIZUTANI, M. C-23 hydroxylation by *Arabidopsis* CYP90C1 and CYP90D1 reveals a novel shortcut in brassinosteroid biosynthesis. **The Plant Cell**, v.18, p.3275-3288, 2006.

PFALZ, M.; VOGEL, H.; KROYMANN, J. The gene controlling the indole glucosinolate modifier1 quantitative trait locus alters indole glucosinolate structures and aphid resistance in *Arabidopsis*. **The Plant Cell**, v.21, p.985-999, 2009.

RALSTON, L.; KWON, S. T.; SCHOENBECK, M.; RALSTON, J.; SCHENK, D. J.; COATES, R. M.; CHAPPELL, J. Cloning, heterologous expression, and functional characterization of 5-epi-aristolochene-1,3-dihydroxylase from *Tobacco (Nicotiana tabacum)*. **Archives of Biochemistry and Biophysics**, v.393, n.2, p.222-235, 2001.

RO, D. K.; MAH, N.; ELLIS, B. E.; DOUGLAS, C. J. Functional characterization and subcellular localization of *Poplar (Populus trichocarpa 3 Populus deltoides)* cinnamate 4-hydroxylase. **Plant Physiology**, v.126, p.317-329, 2001.

SAITO, S.; HIRAI, N.; MATSUMOTO, C.; OHIGASHI, H.; OHTA, S.; SAKATA, K.; MIZUTANI, M. *Arabidopsis* CYP707As encode abscisic acid hydroxylase, a key enzyme in the oxidative catabolism of abscisic acid. **Plant Physiology**, v.134, p.1439-1449, 2004.

SAUVEPLANE, V.; KANDEL, S.; KASTNER, P.; EHLTING, J.; COMPAGNON, V.; WERCK-REICHHART, D.; PINOT, F. *Arabidopsis thaliana* CYP77A4 is the first cytochrome P450 able to catalyze the epoxidation of free fatty acids in plants. **The Federation of European Biochemical Societies Journal**, v.276, p.719-735, 2009

SCHOCH, G.; GOEPFERT, S.; MORANT, M.; HEHN, A.; MEYER, D.; ULLMANN, P.; WERCK-REICHHART, D. CYP98A3 from *Arabidopsis thaliana* is a 3'-hydroxylase of phenolic esters, a missing link in the phenylpropanoid pathway. **Journal of Biological Chemistry**, v.276, p.36566-36574, 2001.

SCHOENBOHM, C., S. MARTENS, C.; EDER, G.; FORKMANN, G., WEISSHAAR, B.; Identification of the *Arabidopsis thaliana* flavonoid 3'-hydroxylase gene and functional expression of the encoded P450 enzyme. **Biological Chemistry**, v.381, N.8, p.11749-11753, 2000.

SCHRÖDER, G.; UNTERBUSCH, E.; KALTENBACH, M.; SCHIDT, J.; STRACK, D.; De LUCA, V.; SCHÖDER, J. Light-induced cytochrome P450-dependent enzyme in indole alkaloid biosynthesis: tabersonine 16-hydroxylase. **The Federation of European Biochemical Societies Journal Letters**, v.458, p.97-102, 1999.

SIMINSZKI, B.; CORBIN, F. T.; WARD, E. R.; FLEISCHMANN, J.; DEWEY, R. Expression of a soybean cytochrome P450 monooxygenase cDNA in yeast and tobacco enhances the metabolism of phenylurea herbicides. **Proceedings of the National Academy of Sciences of the United States of America**, v.96, n.4, p.1750-1755, 1999.

STUMPE, M; FEUSSNER, I. Formation of oxylipins by CYP74 enzymes. **Phytochemistry Reviews**, v.5, p.347-357, 2006.

SWAMINATHAN, S.; MORRONE, S.; WANG, Q.; FULTON, D. B.; PETERS, R. J. CYP76M7 is an ent-cassadiene C11a-hydroxylase defining a second multifunctional diterpenoid biosynthetic gene cluster in Rice. **The Plant Cell**, v.21, p.3315-3325, 2009.

TAMURA, K; PETERSON, D.; PETERSON, N.; STECHER, G.; NEI, M.; KUMAR, S. MEGA5: Molecular evolutionary genetics analysis using likelihood, distance, and parsimony methods. **Molecular Biology and Evolution**, v.10, n.7, p.A13, 2011.

TIAN, L.; MUSETTI, V.; KIM, J.; MAGALLANES-LUNDBACK, M.; DELLAPENNA, D. The *Arabidopsis* LUT1 locus encodes a member of the cytochrome P450 family that is required for carotenoid e-ring hydroxylation activity. **Proceedings of the National Academy of Sciences of the United States of America**, n.101, n.1, p.402-407, 2004.

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**, v.22, n. 22, p.4673-4680, 1994.

WANG, E.; WANG, R.; DEPARASIS, J.; LOUGHRIN, J.H.; GAN, S.; WAGNER, G.J. Suppression of a P450 hydroxylase gene in plant trichome glands enhances natural-product-based aphid resistance. **Nature Biotechnology**, v.19, n.4, p.371–374, 2001.

WANG, Q.; HILLWIG, M. L.; WU, Y. PETERS, R. J. CYP701A8: A rice ent-kaurene oxidase paralog diverted to more specialized diterpenoid metabolism. **Plant Physiology**, v.158, n.1, 2012.

ZHOU, N.; TOOTLE, T. L.; GLAZEBROOK, J. *Arabidopsis* PAD3, a gene required for camalexin biosynthesis, encodes a putative cytochrome P450 monooxygenase. **The Plant Cell**, v.11, p.2419-2428, 1999.

ZHU, Y.; NOMURA, T.; XU, Y.; ZHANG, Y.; PENG, Y.; MAO, B.; HANADA, A.; ZHOU, H.; WANG, R.; LI, P.; ZHU, X.; MANDER, L. N.; KAMIYA, Y.; YAMAGUCHI, S.; HE, Z. Elongated uppermost internode encodes a cytochrome P450 monooxygenase that epoxidizes gibberellins in a novel deactivation reaction in Rice. **The Plant Cell**, v.18, p.442-456, 2006.

## 5 ARTIGO 2 - BIOSÍNTESE DE DITERPENOS EM *COFFEA ARABICA*: CARACTERIZAÇÃO BIOQUÍMICA E ANÁLISE DO PERFIL TRANSCRICIONAL DE GENES CANDIDATOS

### 5.1 INTRODUÇÃO

Os diterpenos são produtos do metabolismo secundário das plantas e são provenientes da síntese dos isoprenóides (WILLE et al., 2004). No gênero *Coffea* foram identificados até o momento quatro tipos de diterpenos e dentro destes o cafestol (Caf) e o caveol (Cav) são os principais (SPEER & KOLLING-SPEER, 2006). Os diterpenos são formados a partir da molécula de geranylgeranyl difosfato (GGPP) para formar o esqueleto caurano de vinte carbonos (WILLE et al, 2004; ROBERTS, 2007).

O Caf é encontrado nos frutos de *C. canephora* e *C. arabica* e somente em folhas de *C. arábica*. E o Cav, por sua vez, é encontrado apenas nos frutos de *C. arabica* (DIAS et al., 2010). O fruto de café possui um perisperma abundante durante as primeiras fases do seu desenvolvimento que decresce com a maturação dando lugar ao endosperma (DE CASTRO & MARRACCINI, 2006). A produção destes diterpenos ocorre principalmente no perisperma e ao longo do desenvolvimento do fruto esses diterpenos são transferidos para o endosperma, similarmente ao que ocorre para o metabolismo da sacarose (GEROMEL et al., 2006; DIAS, et al, 2010; SRIDEVI et al, 2010).

Os diterpenos têm importância no sabor e aroma da bebida (SPEER et al., 1993). Apresentam propriedades antioxidantes e anticarcinogênicas, sendo utilizados na composição de filtros solares (GROLLIER & PLESSIS, 1998; CAVIN et al., 2002; LEE & JEONG, 2007; HIGGINS et al., 2008; HUBER et al., 2008). Por outro lado algumas pesquisas observaram que o Caf pode causar o aumento do nível de colesterol em humanos através da ingestão da bebida (URGERT et al., 1995; POSTE et al., 2000; NAIDOO et al., 2011).

Além das diferenças entre os teores de diterpenos genótipo-específicas e espécie-específicas, a quantidade final de diterpenos pode variar no produto final de acordo com o modo de preparo da bebida: em bebida filtrada foi detectado 0.08mg, em espresso 1.2mg, em escandinavo 2.4mg, em *french press*

2.8mg e em preparação turca pode chegar a 3.12mg em cada xícara contendo 120mL da bebida (NAIDOO et al., 2011).

Pouco se conhece sobre as enzimas/genes que atuam nas etapas finais da biossíntese de Caf e Cav, evidenciando a necessidade da melhor compreensão do funcionamento desta rota de biossíntese. Alguns trabalhos relatam que nas vias de biossíntese de ácidos graxos há a ação de genes de Cyt P450 para catalisar a produção desses compostos (PINOT et al., 2007). Em arroz, WANG et al. (2012), demonstraram que a Cyt P450 CYP701A8 é uma caureno oxidase que realiza hidroxilação na via de biossíntese de giberelinas, convertendo o composto ent-caureno em ácido iso-caurenóico.

Com base nestas evidências já observadas para as Cyt P450s monoxigenases, em plantas é provável que a ação destes genes possa interferir no acúmulo e/ou degradação de Caf e Cav. O objetivo deste estudo visou quantificar Caf e Cav em diferentes fases do desenvolvimento de frutos de café e correlacionar estes dados com a atividade transcricional de genes de Cyt P450 de cafeeiro.

## 5.2 MATERIAIS E MÉTODOS

### 5.2.1 Material Vegetal

Foram utilizados folhas, botões florais, flores e perisperma de frutos de nove plantas adultas e saudáveis de *C. arabica* cv. IAPAR59 com aproximadamente 20 anos de idade, plantadas no campo experimental do Instituto Agrônomo do Paraná, Londrina, Paraná, Brasil. As coletas foram feitas mensalmente após a florada, ocorrida em 20 de setembro de 2010.

Para análise de diterpenos durante o desenvolvimento dos frutos foram utilizadas amostras da safra de 2010/2011 (botões florais, flor, e perisperma dos frutos de outubro, novembro, dezembro, janeiro, fevereiro, março e abril, onde os meses representam as coletas feitas em 30 à 210 DAF, respectivamente). Os tecidos foram imediatamente congelados em nitrogênio líquido e armazenados em freezer -80°C para posterior extração de diterpenos e RNA total.

Também foi realizado um ensaio com a utilização do metil jasmonato (MJ), um indutor do metabolismo secundário em plantas. Foram coletados frutos com 120DAF da safra 2009/2010, coletadas em 8 de fevereiro de 2010. Os frutos

foram cortados junto ao pedúnculo, de maneira semelhante ao realizado por GEROMEL et al., (2006), de modo a expor o perisperma e permitir a infiltração das soluções e foram mergulhados, por 5 minutos em câmara de vácuo, em solução contendo 2mM de MJ ou em água milliQ autoclavada (C). Após a retirada do vácuo os frutos foram deixados em ambiente com luminosidade e temperatura controlada (24 a 26° C). As amostras foram coletadas durante 0, 24, 48 e 96 horas de tratamento, onde parte do material foi utilizada para análise por Cromatografia Líquida de Alta Precisão (CLAE) e parte armazenada em freezer -80°C para extração de RNA.

### 5.2.2 Análise das Concentrações de Cafestol e Caveol por CLAE

Os tecidos do grão de café das amostras foram separados em polpa, perisperma e endosperma e selecionou-se apenas o perisperma para as análises por CLAE (Figura 7), baseado nos estágios de desenvolvimento do fruto (DE CASTRO & MARRACINI et al., 2006) e nas análises de diterpenos feitas previamente por POT et al. (2008) e DIAS et al. (2010).

**Figura 7** - Perisperma do fruto de cafeeiro (novembro, 60DAF).



Imediatamente após a separação de tecidos, estes foram congelados em nitrogênio líquido (para evitar oxidação), pulverizados (com almofariz e pistilo) e armazenados a -80°C.

A quantificação simultânea de cafestol e caveol foi realizada de acordo com a metodologia de CLAE de fase reversa desenvolvida por DIAS et al. (2010) em um Cromatógrafo Surveyor Plus (San Jose, USA) que consiste em um amostrador automático com controle de temperatura Peltier para rack de amostras e forno de colunas integrado (Surveyor PLUS), bomba gradiente (Surveyor LC PLUS),

detector com arranjo de diodos (Surveyor PDA PLUS), cela Lightpipe de 10 $\mu$ L e caminho ótico de 50mm, sistema de dados para cromatografia ChromQuest 5.0 e interface SS420. Foi utilizada uma coluna de fase reversa Spherisorb ODS 1 (250mm x 4.6mm i.d. 5  $\mu$ m), com fase móvel e eluição isocrática de acetonitrila/água (55:45) na vazão de 0,9 mL/min. A detecção foi feita a 220 e 290 nm, para cafestol e caveol, respectivamente. Com temperatura do forno de 25°C, e tempo de corrida de 20 minutos.

A quantificação foi realizada por padronização externa, construindo-se curva de calibração para caveol e cafestol entre 50 a 1000 mg/100g. A extração dos diterpenos foi realizada por saponificação direta com posterior separação da fração insaponificável e utilização cerca de 0,2g de material pulverizado (botão floral, flor ou perisperma dos frutos). Para a saponificação das amostras, utilizou-se hidróxido de potássio em etanol, depois as amostras foram colocadas em banho maria à 80°C por uma hora e limpas com água destilada. Em seguida, realizou-se a extração com terc-butilmetil e limpeza com água purificada (3 vezes) e as amostras foram deixadas em banho maria por 60°C. O extrato etéreo resultante foi ressuspenso na fase móvel composta por acetonitrila e água (diluição 55:45, respectivamente), seguido de diluição do extrato com acetonitrila 100% (1:4, respectivamente), filtração em membrana de nylon 0,45  $\mu$ m (Millipore) e injeção no cromatógrafo.

Foram analisadas 9 plantas, com duas repetições técnicas de extração de Caf e Cav, além de duas injeções de cada amostra extraída, totalizando 36 dados para cada tecido analisado, com exceção de botão floral e flor, no qual foi utilizado material proveniente de um *bulk* de 7 plantas diferentes, com duas repetições técnicas e duas injeções de cada extração, totalizando 4 dados para cada tecido.

As médias e desvios padrões foram determinados pelo programa Statistica versão 6.0 (<http://www.statsoft.com.br>), o qual realizou também as análises de variância (ANOVA) e teste das médias (Tukey) com 9 variáveis independentes (tecidos coletados: botão floral, flor e frutos em 30, 60, 90, 120, 150, 180 e 210DAF) com  $p < 0,05$ .

### 5.2.3 Identificação e Validação de Cyt P450 de *Coffea* e Seleção de Genes Candidatos

As análises de identificação e validação de Cyt P450s para a escolha dos genes candidatos foram realizadas de acordo com o protocolo descrito no artigo 1: “A Super Família das Citocromos P450s em *Coffea*”.

### 5.2.4 Desenho de Oligonucleotídeos para qPCR

A partir das análises *in silico*, foram desenhadas pares de oligonucleotídeos para os genes candidatos de Cyt P450 visando a análise do número de transcritos para cada gene por qPCR. Os *primers* foram desenhados com o programa Primer Express 1.0 (incluso no modelo 7500 Fast Real-Time PCR Applied Biosystem) com *amplicon* entre 80 e 100 pares de bases, temperaturas de *melting* entre 58°C e 60°C e porcentagens de GC entre 30 e 60%.

### 5.2.5 Extração de RNA

RNA total dos tecidos foi extraído seguindo o protocolo de CHANG et al. (1993). A quantificação do RNA foi feita com equipamento NanoDrop 1000 Spectrophotometer da Thermo Scientific. Para verificar a qualidade do RNA extraído, as amostras foram submetidas a eletroforese com tampão TAE 1x em gel de agarose 1% e visualizados após coloração com brometo de etídio (0,5 µg/ml) sob luz ultravioleta. Para verificar a contaminação com DNA genômico foram feitos PCRs das amostras de RNA com o gene GAPDH (gliceraldeído-3-fosfato desidrogenase) com base nos primers indicados por CRUZ et al. (2009) e utilizado como controle positivo o DNA de *C. arabica* cv. IAPAR 59.

### 5.2.6 Análise Transcricional por qPCR

A expressão dos genes selecionados foi detectada através de PCR em Tempo Real (aparelho 7500 Fast Real-Time PCR System, Applied Biosystems) entre tecidos contrastantes para os teores de diterpenos (alta e baixa concentração

de Caf e Cav). Foram utilizadas folhas, flores e perisperma (90, 120 e 150 DAF) de *Coffea arabica* cv. IAPAR59.

Para isso, 5 $\mu$ g de RNA total de cada amostra biológica foi utilizado para síntese de cDNA com o kit Superscript<sup>TM</sup> III Reverse Transcriptase (INVITROGEN), conforme especificações do fabricante. As reações de quantificações relativas por PCR em tempo real (qPCR) foram feitas com o kit PCR Master Mix SYBR Green da Applied Biosystems, conforme recomendação do fabricante. Para cada reação foram utilizados 12,5 $\mu$ l de SYBR Green, 0,5 $\mu$ l de primer forward e 0,5 $\mu$ l de primer reverse, 1  $\mu$ l de cDNA, 10,5  $\mu$ l de água MilliQ autoclavada, totalizando um volume final de 25  $\mu$ l.

Quantidades equivalentes de cDNA foram utilizadas para a reação de PCR utilizando-se os pares de *primers* específicos para cada gene de interesse. Estes *primers* foram avaliados conforme as suas eficiências de amplificação *in silico* com o programa LinRegPCR (RAMAKERS et al., 2003). Foi utilizada diluições seriadas dos cDNAs de  $10^0$ ,  $10^{-1}$ ,  $10^{-2}$ ,  $10^{-3}$ , e foi utilizada a diluição de  $10^{-1}$ , a qual apresentou maior estabilidade para os tecidos analisados.

Foram efetuadas três repetições analíticas para cada gene, nos diferentes tecidos e repetições em triplicata biológica, para tecidos contrastantes para o acúmulo/degradação de Caf e Cav. Os normalizadores GAPDH e Ubiquitina10 (UBQ10) foram selecionados a partir de trabalhos anteriores em cafeeiros (BARSALOBRES-CAVALLARI et al., 2009; CRUZ et al., 2009). No teste de eficiência dos oligonucleotídeos a expressão de UBQ10 foi considerada mais estável e por isso este gene foi utilizado para os cálculos de normalização da expressão relativa e como calibrador foi utilizado o padrão de atividade transcricional das folhas.

O  $\Delta C$  foi calculado para cada tipo de amostra/oligonucleotídeo e foi utilizado como calibrador o tecido folha. O cálculo de expressão relativa (RQ) foi feito de acordo com a metodologia desenvolvida por LIVAK & SCHMITTGEN (2001) através da fórmula:  $RQ = 2^{-\Delta\Delta C_t}$ , onde  $\Delta\Delta C_t = \Delta C_t(\text{amostra}) - \Delta C_t(\text{calibrador})$ .

### 5.3 RESULTADOS E DISCUSSÃO

#### 5.3.1 Quantificação de Cafestol e Caveol por CLAE

A quantificação simultânea de Caf e Cav em botão floral, flor e perisperma dos frutos de cafeeiros das amostras da safra de 2010/2011 (novembro, dezembro, janeiro, fevereiro, março, abril e maio) e de frutos de fevereiro (120 DAF), tratados com metil jasmonato (24, 48 e 96 horas) ou H<sub>2</sub>O (24, 48 e 96 horas) foi feita por CLAE e estão apresentadas na tabela 8 e 9, e nas figuras 8 e 9 (anexo 5).

**Tabela 8** - Concentrações médias de Caf e Cav ao longo do desenvolvimento do fruto (safra 2010/2011)

Amostras	Média ± Desvio Padrão	
	Caf (mg/100g)	Cav (mg/100g)
Botão Floral	308.03 ± 9.7 a	11.42 ± 0.8 de
Flor	295.53 ± 20.9 a	10.09 ± 0.3 de
Outubro (30DAF)	241.68 ± 44.5 bc	81.77 ± 41.2 de
Novembro (60DAF)	217.98 ± 20.9 cd	304.11 ± 47.5 c
Dezembro (90DAF)	203.16 ± 37.1 d	513.57 ± 85.6 b
Janeiro (120DAF)	276.67 ± 29.9 ab	1009.48 ± 71.3 a
Fevereiro (150DAF)	69.83 ± 17.5 e	355.24 ± 85.1 d
Março (180DAF)	24.46 ± 10.2 f	121.48 ± 70 e
Abril (210DAF)	17.93 ± 2.6 f	14 ± 5.7 e

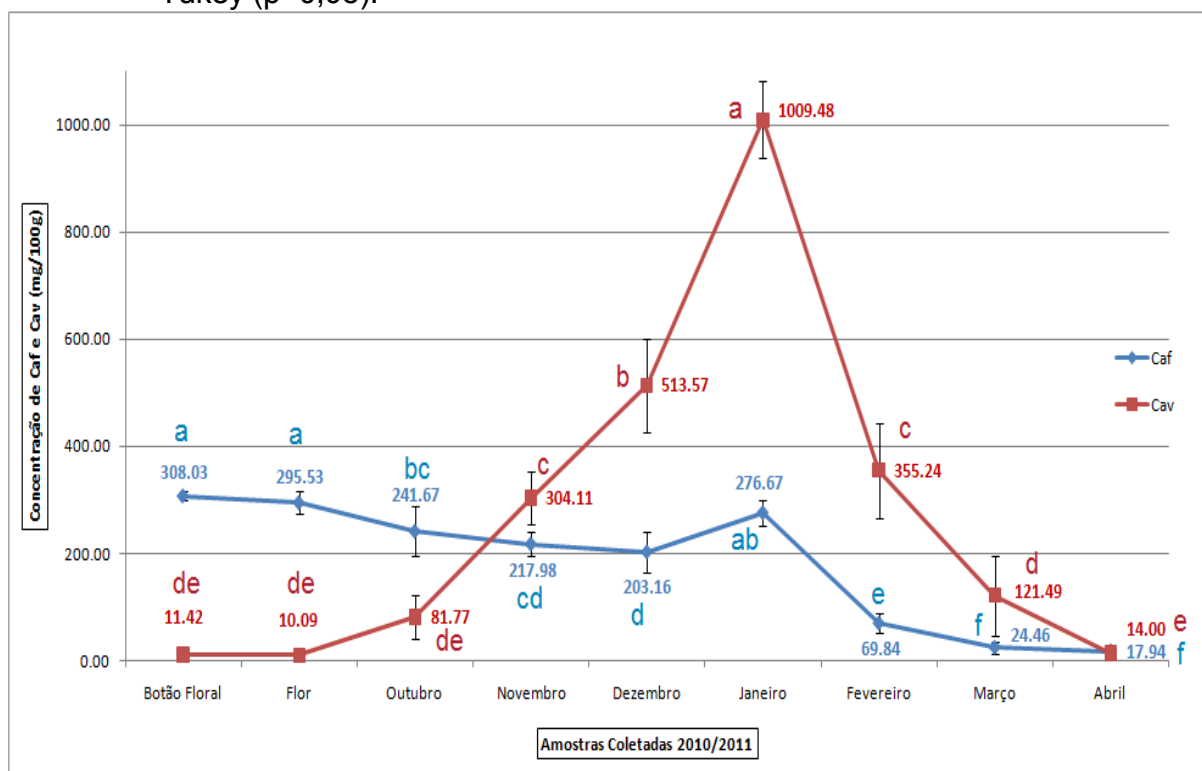
Os dados são referentes à média das concentrações das amostras. DAF = dias após florada. Médias seguidas de mesma letra na coluna não diferem entre si pelo teste de Tukey ( $p < 0,05$ ).

Baseado nos dados da tabela 8 e na figura 8 observa-se um aumento gradativo de Cav o qual atinge sua quantificação máxima em janeiro com uma redução acentuada nos meses seguintes. Para Caf, inicialmente se observa uma alta concentração em botões florais e durante o desenvolvimento da flor e do desenvolvimento inicial do perisperma essa concentração diminui, aumentando em janeiro (pico em 120DAF), e com queda gradativa até abril (210DAF). Um padrão similar foi observado para o Cav após o período de maior acúmulo em 120DAF, com a diminuição da concentração dos diterpenos concomitante com a maturação e formação do grão.

A maior diferença média entre as amostras foi detectada no mês de janeiro com relação a dezembro onde houve um aumento de 284% para os teores de Cav. Também foi observada uma diferença significativa no mês de janeiro com

relação a fevereiro onde houve um decréscimo de 396% para os teores de Caf, como demonstrado na figura 8. Estes dados corroboram com os encontrados por POT et al (2008) e DIAS et al. (2010), cujos resultados foram similares e os maiores teores de Caf e Cav foram detectados entre 90 e 120DAF (dezembro e janeiro, respectivamente).

**Figura 8** - Níveis médios de Caf e Cav por CLAE de botão floral, flor e do perisperma de frutos ao longo do seu desenvolvimento. Cav (azul) e Caf (vermelho). Médias seguidas de mesma letra na linha não diferem entre si pelo teste de Tukey ( $p < 0,05$ ).



Na análise por CLAE, foi observada uma inversão nas concentrações de Caf e Cav de botão floral e flores com relação aos frutos. A quantificação de diterpenos ao longo do desenvolvimento do perisperma apresentou altos teores, crescentes até cerca de 120 DAF. Segundo DE CASTRO & MARRACCINI (2006), a partir deste período, o perisperma vai dando espaço para o desenvolvimento do endosperma, fator pelo qual surge a hipótese de que os diterpenos são transferidos para este último tecido nas fases finais de maturação do fruto. A maioria dos estudos de grãos de café realiza as medições nos frutos maduros ou com grãos torrados como é o caso de SRIDEVI et al., (2010) o que não

havia permitido elucidar até o momento a dinâmica de acúmulo destes diterpenos durante a formação do fruto.

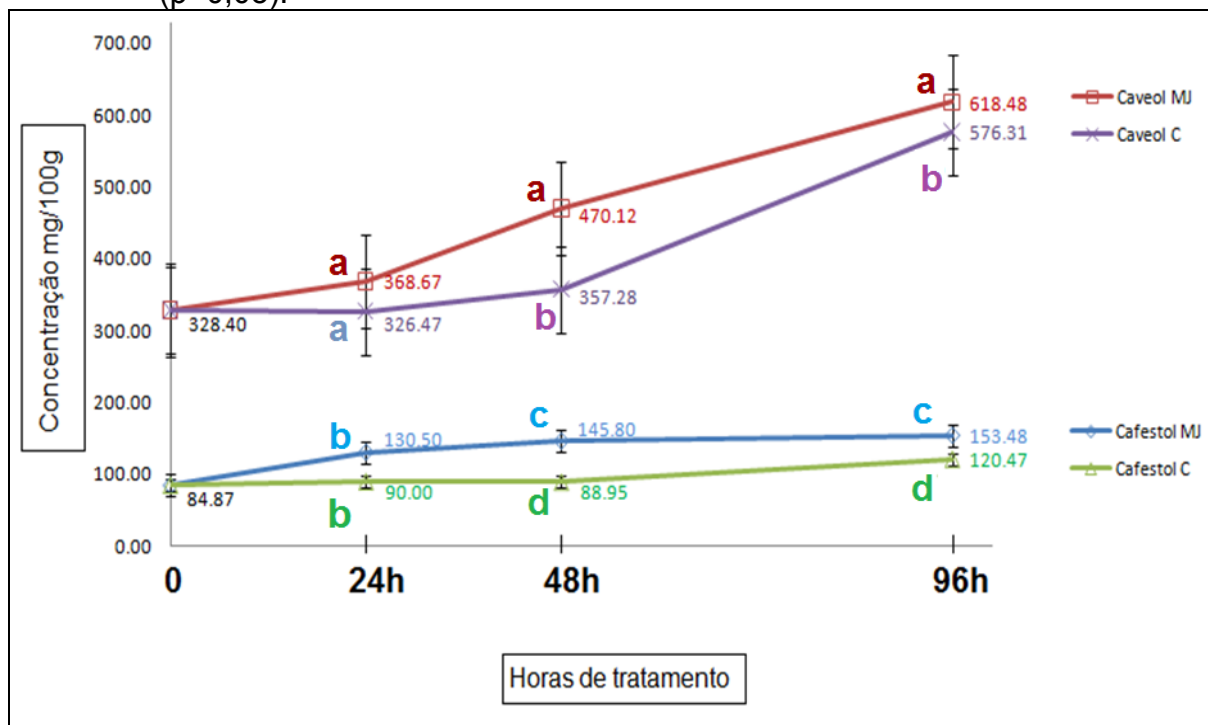
Com relação à aplicação de metiljasmonato em frutos com 120DAF foi possível observar um aumento de Caf (31,03%) e de Cav (11,44%) em relação ao controle 24 horas após aplicação (MJ1). Essa diferença foi mais acentuada 48 horas após aplicação, com uma diferença de 38,99% de Caf e 24% de Cav em relação ao controle (figura 9 e tabela 9).

**Tabela 9** - Concentrações médias de diterpenos em frutos tratados com MJ (safra 2009/2010)

Amostras	Média ± Desvio Padrão	
	Caf (mg/100g)	Cav (mg/100g)
C (24h)	90 ± 8 b	326.47 ± 27.8 a
C (48h)	88.95 ± 3.4 d	357.88 ± 30.2 b
C (96h)	120.47 ± 7.7 d	576.31 ± 15.9 b
MJ (24h)	130.5 ± 7.79 b	368.67 ± 19.0 a
MJ (48h)	145.8 ± 3.3 c	470.12 ± 16.3 a
MJ (96h)	153.48 ± 8.7 c	618.48 ± 14.2 a

C= frutos inoculados com água (controle); MJ = amostras tratadas com metil jasmonato. Médias seguidas de mesma letra na coluna não diferem entre si pelo teste de Tukey ( $p < 0,05$ ).

**Figura 9** - Níveis médios de Caf e Cav por CLAE nas amostras controle e com aplicação de metil jasmonato (24, 48 e 96 horas) da safra de 2009/2010 (fevereiro, 120DAF). Cav (MJ:vermelho; C: roxo) e Caf (MJ: azul; C: verde). Médias seguidas de mesma letra não diferem entre si pelo teste de Tukey ( $p < 0,05$ ).



Com isto, pode-se verificar que a utilização do indutor de metabolismo secundário MJ na concentração de 2mM foi eficiente para aumentar os níveis de Caf e Cav nos frutos. Em *Nigella sativa* a utilização de 100 $\mu$ M, de MJ em semente em cultivo hidropônico aumentou em 12 vezes os triterpenóides  $\alpha$ -hederin e kalopanaxsaponin (Ksl), e esta via metabólica é catalisada por uma Cyt P450 que transforma o composto  $\beta$ -amyrin em  $\alpha$ -hederin que posteriormente dá origem a Ksl (SCHOLZ et al., 2009).

### 5.3.2 Seleção e Análise dos Genes Candidatos para a Biossíntese de Caf e Cav

Do total de 84 contigs identificados nas análises de identificação e validação de Cyt P450s em cafeeiro, 14 contigs foram pré selecionados (tabela 10), os contigs 19, 22, 23, 25, 28, 33, 37, 41, 58, 80, 96, 100, 102 e 107, pois eram constituídos por sequências provenientes de bibliotecas de frutos, tecido no qual os níveis de Caf e Cav são muito maiores quando comparados as folhas (DIAS et al., 2010).

**Tabela 10** - Resultado da Seleção dos Genes Candidatos

Tipo P450	Contigs Novos	No Reads Total	No Reads Frutos	Família	Tamanho Proteína	pb
A	<b><i>Contig 19</i></b>	<b>18</b>	<b>9</b>	<b>CYP76</b>	<b><i>F.L.=471aa</i></b>	<b>1560</b>
	<b><i>Contig 22</i></b>	<b>16</b>	<b>10</b>	<b>CYP81</b>	<b><i>229 a 511aa(511)</i></b>	<b>1073</b>
	<b><i>Contig 23</i></b>	<b>15</b>	<b>14</b>	<b>CYP94</b>	<b><i>169 a 498(498)</i></b>	<b>1270</b>
	<b><i>Contig 28</i></b>	<b>13</b>	<b>8</b>	<b>CYP82</b>	<b><i>1 ao 508aa(519)</i></b>	<b>1592</b>
	<b><i>Contig 33</i></b>	<b>11</b>	<b>nd*</b>	<b>CYP71</b>	<b><i>F.L.=513aa</i></b>	<b>1750</b>
	Contig 80	3	3	CYP92	331 ao 509 (509)	762
	Contig 96	2	2	CYP71	529 ao 768 (1701)	771
	Contig 100	2	2	CYP736	307 ao 466 (496)	704
	<b><i>Contig 107</i></b>	<b>22</b>	<b>8</b>	<b>CYP701</b>	<b><i>88 ao 437(437aa)</i></b>	<b>1049</b>
não A	<b><i>Contig 25</i></b>	<b>14</b>	<b>8</b>	<b>CYP72</b>	<b><i>F.L.=519aa</i></b>	<b>1943</b>
	<b><i>Contig 37</i></b>	<b>10</b>	<b>10</b>	<b>CYP74</b>	<b><i>F.L. = 488aa</i></b>	<b>1975</b>
	Contig 41	8	8	CYP714	1 ao 462 (524)	1544
	Contig 58	5	5	CYP96	1 ao 242 (502)	1010
	Contig 102	2	2	CYP94	1 ao 155 (507)	545

Contigs em **negrito** e *itálico* são os genes candidatos selecionados por serem quase todos completos para a codificação da proteína (F.L.=*full lengths*, ou seja, codificação completa). Nd\* = não disponível, porém o contig 33 apresentou 7 reads relacionados a estresses abióticos.

A segunda pré seleção foi feita na escolha dos contigs com um número maior de sequências, para aumentar a eficiência no desenho de *primers* para os mesmos. Desta forma, de 14 contigs, foram selecionados apenas oito contigs (19, 22, 23, 25, 28, 33, 37 e 107).

Depois de selecionados os oito genes candidatos para análises de qPCR esses passaram por uma análise de Blast2Seq (*blast two sequences*) pelo software CLC workbench para eliminar a possibilidade de serem isoformas iguais.

As isoformas foram confirmadas como sendo diferentes entre si, de acordo com a porcentagem de identidade encontrada que variou entre 16,82 e 45,47% (anexo 6). Para a superfamília das aquaporinas um *cutoff* de 95% de identidade entre os aminoácidos foi utilizada para separação das isoformas (JAVOT et al., 2003).

As sequências consenso das oito Cyt P450s foram alinhadas pelo software CLC workbench para verificar as suas similaridades, localização das ORFs (regiões codificadoras de proteínas), e regiões distintas, com o objetivo de desenhar *primers* evitando locais de domínios conservados ou com alto índice de similaridade.

Os oligonucleotídios foram testados por PCR normal através da amplificação de DNA de *Coffea arabica* (cv. IAPAR59) e baseado neste teste, os

*primers* foram selecionados para análise de atividade transcricional por qPCR foram os contigs C19\_2, C22\_1, C23\_1, C25\_2, C28\_1, C33\_1, C37\_2 e 107\_1, por apresentarem amplificação específica em torno de 80 à 100pb (tabela 11). As eficiências de amplificação de cada *primer* também foi calculada pelo programa LinRegPCR.

**Tabela 11** - Lista dos *Primers* desenhados para os Genes Candidatos

<i>Contigs</i>	<i>Primer forward</i>	<i>Primer reverse</i>	Eficiência	Tam Amplicon
C19_2	GAAGAACTGGGGAAGCCCT	AGAACGTGGTCATGGTCTGAATTC	99,5%	100pb
C22_1	GCTGTCTTTGCAAGCCTGG	GAAGAACTGGGGAAGCCCT	99,75%	99pb
C23_1	ACTCCGGACCCCTCTATGTTG	GCCCTGGGATTTTCTCGAT	97,4%	100pb
C25_1	ACTCCGGACCCCTCTATGTTG	CCCTGGGATTTTCTCGACC	98,35%	100pb
C28_1	TGTCCAGAAATCCAAGTTGGTG	GGACAGTAATTCTCCAAGTCTTGTG	97,4%	100pb
C33_1	CCAGAAATCCAAGTTGGTGATT	GGACAGTAATTCTCCAAGTCTTGTG	99,45%	100pb
C37_2	AGGCTGTTGGGAAAGTCTTC	ACTGTTGGAAGTCTCGAATGC	91,9%	103pb
C107_1	TCGGGCTTACCAACCAAAAA	GATGTGGTCGAGGCTTTCCA	110%	80pb

### 5.3.3 Perfil Transcricional dos Genes Candidatos de Cyt P450

A expressão relativa das 8 Cyt P450s foram em suma maioria diferentes entre si, indicando que se tratam de isoformas de subfamílias diferentes, o que corrobora com as análises *in silico* (figuras 10 e 11).

Dois dos genes selecionados, C22\_1 e C25, apresentaram um padrão de expressão maior em folhas do que nos outros tecidos, onde não houve diferença de expressão (figura 10-B e 11-A). Esses dois genes possivelmente não estão envolvidos na via de diterpenos, mas com a biossíntese de outros metabólitos. A isoforma C22\_1 é uma putativa CYP81 da família tipo A, relacionada com a biossíntese de glicosinolatos indólicos (iGS), convertendo I3M (indole-3-metil glucosinolato) em 4MOI3M (4-metoxi-indole-3-metil) via 4-hidroxi-indole-3-metil glucosinolato (4HOI3M) (PFALZ et al., 2009; KAI et al., 2011). Estes compostos estão ligados a mecanismos de proteção da planta contra estresses bióticos (herbívoros e microrganismos patógenos) e são considerados importantes para o crescimento e desenvolvimento da planta. O contig C25\_1 é uma putativa CYP72 da família tipo não A, e segundo IMAISHI & MATUMOTO, (2007) as CYP72A18 estão relacionados a detoxificação da planta ao ácido herbicida palargônico em sementes de arroz e à sinalização de moléculas de defesa da planta, além disto GUTTIKONDA

et al. (2010) observou uma maior expressão da CYP72A135 em folhas (10 vezes mais do que em flores) de soja com dados de sequenciamento em larga escala de transcritos. Resultados similares foram encontrados por XU et al. (2001), observaram expressão relativa maior em folhas do que em flores tanto para a CYP72A18 como para a CYP81F3 em *Arabidopsis thaliana*.

O contig C23\_1 da família tipo A apresentou uma expressão baixa em flores e no desenvolvimento do perisperma, com um aumento de expressão aos 150DAF (figura10-C). Este gene é similar as CYP94B3 (família tipo não A) que segundo KOO et al. (2011) estão relacionadas a biossíntese de fitohormônios (jasmonoil-L-isoleucinas) envolvidos no crescimento, desenvolvimento e mecanismos de defesa da planta contra ataques de insetos. Isto justifica a expressão dessa P450 em folhas para sua proteção contra insetos e nos frutos pode estar sendo novamente requerida aos 120 DAF para o crescimento e desenvolvimento pois é nessa fase em que e o endosperma começa a se desenvolver, ocupando o espaço do perisperma.

As isoformas C19\_2 (figura 10-A) e C37\_2 (figura 11-B) apresentaram um padrão transcricional semelhante entre si e com uma alta atividade transcricional em flores, baixa atividade relativa em folhas e em frutos com 90DAF, os quais apresentam um aumento gradativo de expressão aos 120 até 150DAF.

O contig C19\_2 é uma putativa CYP76 do tipo A, nesta subfamília GUTTIKONDA et al. (2010) encontraram em *Glycine max* uma alta expressão da isoforma CYP76F17 em flores em comparação com os demais tecidos. Essa subfamília também está relacionada à biossíntese de alcalóides e mecanismos de defesa, como por exemplo, na detoxificação de herbicidas por enzimas produzidas pelas plantas e na produção de diterpenos (fitoalexinas) que atuam no combate à infecções fúngicas e bacterianas em *Oryza sativa* (DIDIÉRIE, et al., 2002; SWAMINATHAN et al, 2009). Em *A. thaliana* foi observado uma expressão quatro vezes maior da CYP76C3 em flor do que nas folhas (XU et al., 2001).

O contig C37\_2 apresenta similaridade com as CYP74 do tipo não A que segundo STUMPE & FEUSSNER (2006) catalisam reações para a formação de compostos aldeídos, álcoois voláteis e hidroxilação de ácidos graxos. Ela também foi descrita como uma aleno óxido sintase (AOS) e uma hidroperoxiliase (HPL), que atuam na formação de hormônios de defesa da planta (ácido jasmônico) e na

formação de compostos voláteis que podem atrair inimigos naturais de insetos herbívoros (GUTTIKONDA et al, 2010). Um trabalho realizado por XU et al. (2001) também detectou uma grande diferença de expressão das CYP74B2 em flores através de microarranjo e não foi relatado a detecção em folhas.

A transcrição do contig C28\_1 (figura 10-D) também foi acentuada em flores. A análise em perisperma também apresentou uma expressão relativa maior que de folhas, com um pico em 150DAF. O contig C28\_1 é similar as CYP82, que estão envolvidas na biossíntese de homoterpenos, compostos responsáveis por contribuir no odor floral. Similar as CYP74, estes compostos também estão relacionados com à atração de polinizadores e exercem função de afastar insetos herbívoros, dessa forma contribuindo para a sobrevivência e reprodução da planta (LEE et al., 2010; THOLL et al., 2011).

É interessante observar que a expressão relativa deste contig foi 14 mil vezes mais expresso do que em folhas. Apesar desses valores serem incomuns trabalhos com diferenças desse nível de expressão já foram observadas em *Brassica napus* e *Curcubita maxima* (PANT et al., 2008), em ervilha (FOO et al., 2005) e em células tumorais humanas (OBERTHUER et al., 2004).

Como os diterpenos estão envolvidos na formação de compostos aromáticos de *Coffea*, a alta expressão desses três genes em flores pode estar relacionadas a maior concentração de cafestol encontrada pelas análises de CLAE nas flores de cafeeiros, além da formação de outros compostos voláteis.

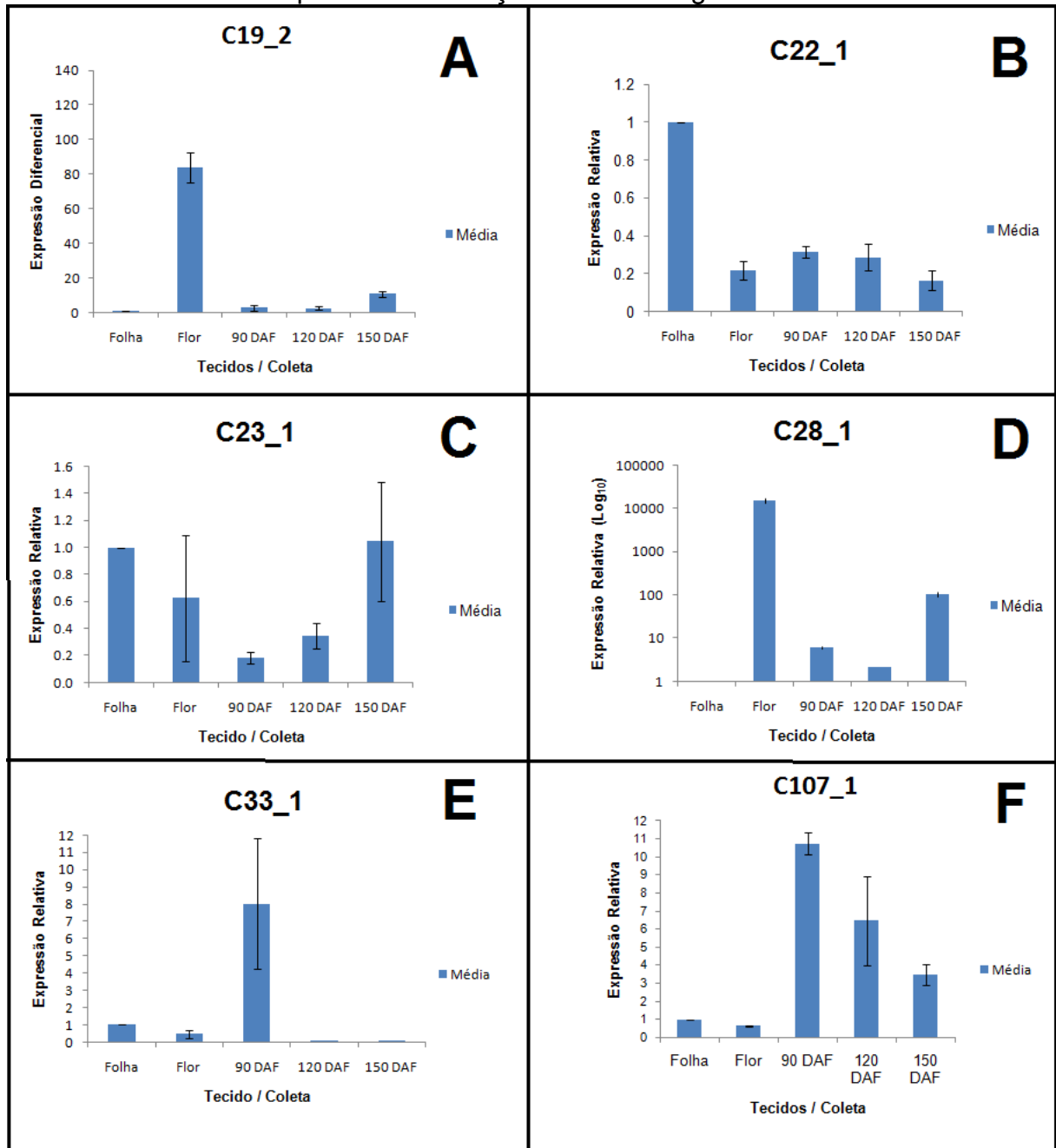
Nos contigs C33\_1 (figura 10-E) e C107\_1 (figura 10-F) houve um pico de expressão em 90DAF, enquanto que no C33\_1 não foi observado diferenças na expressão em outras fases do perisperma, para o C107\_1, foi possível verificar uma expressão relativa maior do que em folhas também aos 120 e 150 DAF. A expressão destes contigs apresentam um padrão esperado para os genes relacionados com a síntese de Caf e Cav. No caso do C107\_1 a linha de tendência acompanhou os dados obtidos por CLAE neste trabalho e descritos por DIAS et al., (2010). Estas Cyt P450s pertencem a família do tipo A e à subfamília das CYP71 (C33\_1) e CYP701 (C107\_1).

A subfamília CYP71 é a que possui o maior número de genes de Cyt P450s identificadas na literatura com inúmeras funções descritas (FREY et al., 1997; BAK et al., 1998; LUPIEN et al., 1999; SCHRÖDER et al., 1999; SIMINSZKY et al., 1999; ZHOU et al., 1999; RALSTON et al., 2001; WANG et al., 2001). De

acordo com estudo de expressão semi-quantitativa realizado por LI et al. (2010) durante o desenvolvimento do fruto, a CYP71A2 de tomate foi altamente expressa no início do desenvolvimento do fruto, enquanto que níveis médios foram detectados nas folhas, flores, botão floral, raízes e caule. O mesmo autor infere que esta Cyt P450 provavelmente está envolvida no desenvolvimento dos frutos e pode estar relacionada a hidroxilação de monoterpenos.

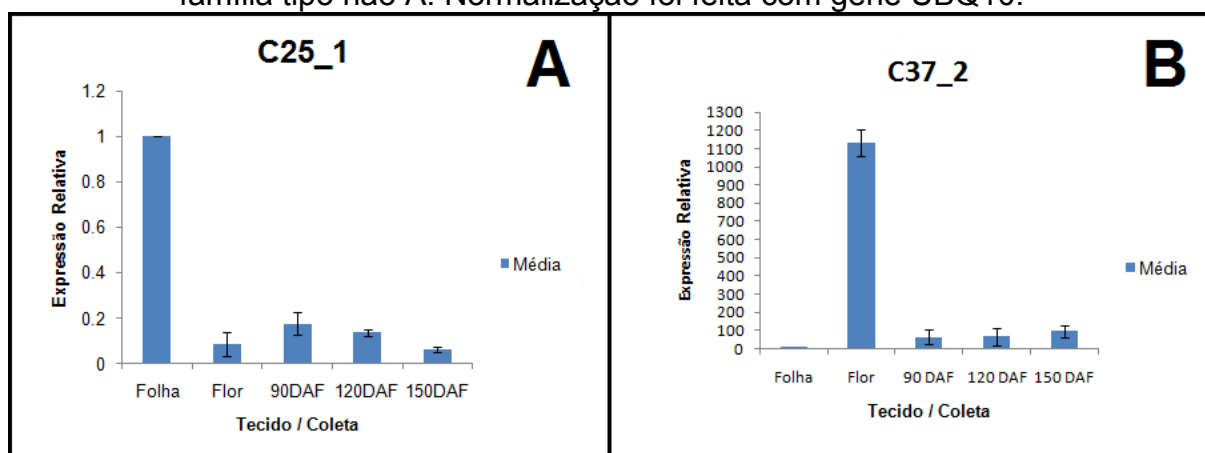
O último gene candidato, C107\_C1, pertence à família do tipo A e à subfamília das CYP701. As Cyt P450 dessa subfamília são descritas como uma ent-caureno oxidase (MORRONE et al., 2011) que realiza três reações de catálise nas etapas iniciais de biossíntese de giberelinas (GA) e brassinoesteróis, e seu último produto (ácido ent-caureno) que é considerado o precursor dos diterpenos Caf e Cav. Não foram encontrados trabalhos até o presente momento que relatem e comparem o seu perfil de expressão em diferentes tecidos.

**Figura 10** - Perfis Transcricionais dos genes candidatos de *Coffea arabica* Cyt P450 da família tipo A. Normalização feita com o gene UBQ10.



Expressão relativa ao padrão encontrado em folha.

**Figura 11** - Perfis Transcricionais de genes candidatos de *C. arabica* Cyt p450 da família tipo não A. Normalização foi feita com gene UBQ10.

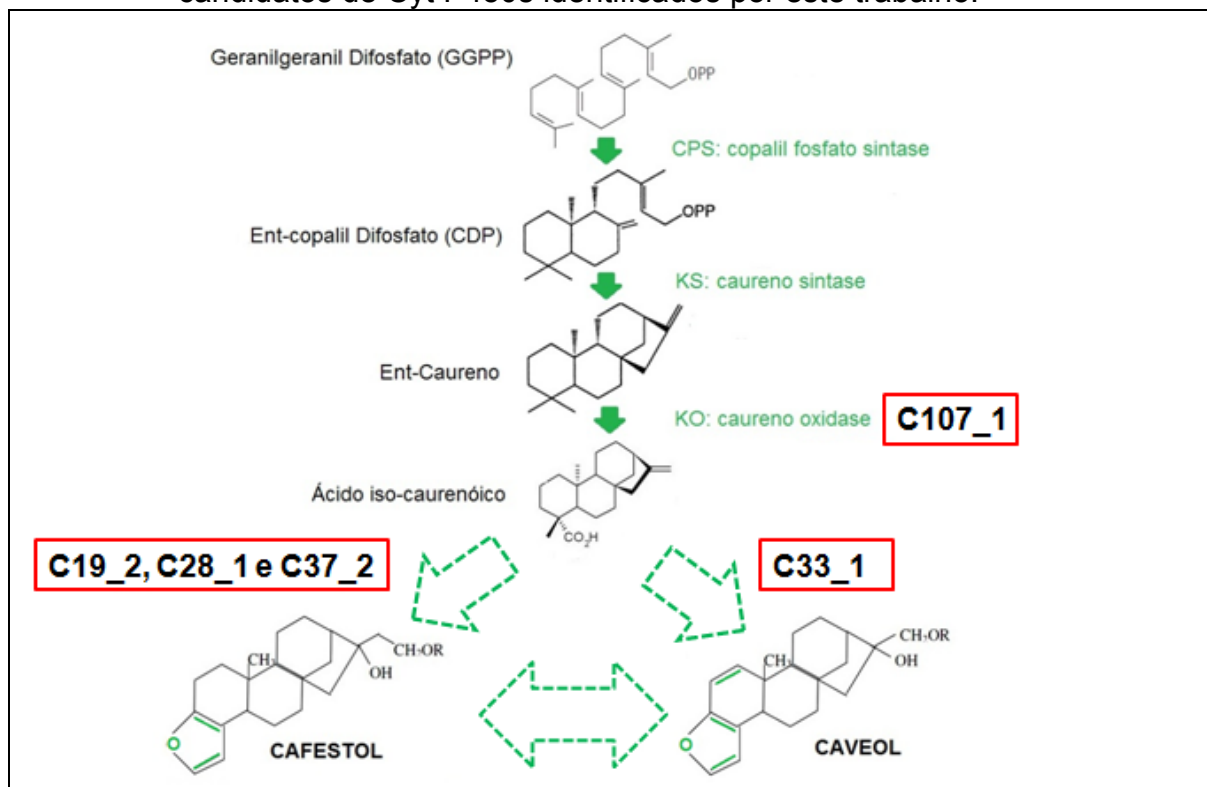


Expressão relativa ao padrão encontrado em folha.

Este estudo além de fornecer informações sobre as concentrações de Caf e Cav em diferentes tecidos e fases do desenvolvimento de *C. arabica* é um dos primeiros trabalhos de análise de genes da super família das Cyt P450s em *Coffea* spp. Os resultados apresentados pela análise de CLAE de flores e desenvolvimento de perisperma associado a análise das atividades transcricionais, permitiram identificar quatro genes envolvidos na formação dos diterpenos Cav e Caf. Os genes dos contigs C33\_1 e C107\_C1, assim como os C19\_1, C\_28 e C37\_2 são fortes candidatos para estudos de caracterização funcional visando identificar as enzimas chaves na formação de Cav e Caf.

De acordo com os resultados obtidos, uma hipótese foi feita para a via de biossíntese de Caf e Cav com a provável atuação dos genes candidatos para estes compostos (figura 12). Estudos de complementação das atividades em sistemas heterólogos, poderão elucidar o papel e as funções destas Cyt P450s em *Coffea arabica*.

**Figura 12** - Hipótese da via de biossíntese de Caf e Cav com a inclusão dos genes candidatos de Cyt P450s identificados por este trabalho.



A longo prazo, estudos de variabilidade genética dos genes identificados poderão ser utilizados em trabalhos de melhoramento genético de mapeamento associativo e como parâmetros para a seleção de novos cultivares com melhor características nutraceuticas dos grãos de café.

#### 5.4 REFERENCIAL BIBLIOGRÁFICO

BAK, S.; OLSEN, C. E.; HALKIER, B. A.; MOLLER, B. L.; Transgenic *Tobacco* and *Arabidopsis* plants expressing the two multifunctional *Sorghum* cytochrome P450 enzymes, CYP79A1 and CYP71E1, are cyanogenic and accumulate metabolites derived from intermediates in dhurrin biosynthesis. **Plant Physiology**, v.123, p.1437-1448, 2000.

BARSALOBRES-CAVALLARI, C. F.; SEVERINO, F. E.; MALUF, M. P.; MAIA, I. G.; Identification of suitable internal control genes for expression studies in *Coffea arabica* under different experimental conditions. **BMC Molecular Biology**, v.10, v.1, 2009.

CAVIN, C.; HOLZHAUSER, D.; CONSTABLE, A.; HUBET, AC.; SCHILTER, B. Cafestol and Kahweol, two specific diterpenes with anticarcinogenic activity. **Food and Chemical Toxicology**, Oxford, v. 40, p. 1155-1163, 2002.

CHANG, S.; PURYEAR J.; CAIRNEY, J. A simple and efficient method for isolating RNA from pine trees. **Plant Molecular Biology**, v. 11, p. 113-116, 1993.

CRUZ, F.; KALAOUN, S.; NOBILE, P., COLOMBO, C.;ALMEIDA, J.; BARROS, L. M. G.; ROMANO, E.; GROSSI-DE-SÁ, M. F.; VASLIN, M.; ALVEZ-FERREIRA, M. Evaluation of coffee reference genes for relative expression studies by quantitative real-time RT-PCR. **Molecular Breeding**, v.23, n.4, p.607-616, 2009.

DE CASTRO, R. D.; MARRACCINI, P. Cytology, biochemistry and molecular changes during coffee fruit development. **Brazilian Journal of Plant Physiology**, v. 18, n.1, 2006.

DIAS, R. C. E.; CAMPANHA, F. G.; VIEIRA, L. G. E.; FERREIRA, L. P.; POT, D.; MARRACCINI, P.; BENASSI, M. T.; Evaluation of kahweol and cafestol in *Coffea* tissues and roasted coffee by a new high-performance liquid chromatography methodology. **Journal of Agricultural and Food Chemistry**, v. 58, n. 1, p. 88-93, 2010.

DIDIERJEAN, L.; GONDET, L.; PERKINS, R.; LAU, S. C.; SCHALLER, H.; O'KEEFE, D. P.; WERCK-REICHHART, D. Engineering herbicide metabolism in *Tobacco* and *Arabidopsis* with CYP76B1, a cytochrome P450 enzyme from Jerusalem Artichoke. **Plant Physiology**, v.130, p.179-189, 2002.

FOO, E.; BULLER, E.; GOUSSOT, M.; FOUCHER, F.; RAMEAU, C.; BEVERIDGE, C. A. The branching gene RAMOSUS1 mediates interactions among two novel signals and auxin in Pea. **The Plant Cell**, v.17, p.464-474, 2005.

FREY, M.; CHOMET, P.; GLAWISCHNIH, E.; STETTNER, C.; GRÜN, S.; WINKLMAIR, A.; WISENREICH, W.; BACHER, A.; MEELEY, R. B.; BRIGGS, S. P.; SIMCOX, K.; GIERL, A.; Analysis of a chemical plant defense mechanism in grasses. **Science**, n.277, p.696-699, 1997.

GEROMEL, C.; FERREIRA, L. P. F.; GUERREIRO, S. M. C.; CAVALARI, A. A.; POT, D.; PEREIRA, L. F. P.; LEROY, T.; VIEIRA, L. G. E.; MAZZAFERA, P. MARRACCINI, P.; Biochemical and genomic analysis of sucrose metabolism during coffee (*Coffea arabica*) fruit development. **Journal of Experimental Botany**, v. 57, n. 12, p.3243-3258, 2006.

GROLLIER, J. F.; PLESSIS, S.; US Patent 4,793,990 **1998**.

GUTTIKONDA, S. K.; TRUPTI, J.; BISHT, N. C.; CHEN, H.; AN, Y. C.; PANDEY, S.; XU, D.; YU, O.; Whole genome co-expression analysis of soybean cytochrome P450 genes identifies nodulation-specific P450 monooxygenases. **BioMed Central Plant Biology**, v. 10, n.243, 2010.

HIGGINS, L. G.; CAVIN, C.; ITOH, K.; YAMAMOTO, M.; HAYES, J. D. Induction of cancer chemopreventive enzymes by coffee is mediated by transcription factor Nrf2. Evidence that the coffee-specific diterpenes cafestol and kahweol confer protection against acrolein. **Toxicology and Applied Pharmacology**, San Diego, v. 226, n. 3, p. 328-337, 2008.

HUBER, W. W.; ROSSMANITH, W.; GRUSCH, M.; HASLINGER, E.; PRUSTOMERSKY, S.; PETER-VÖTÖSMARTY, B.; PARZEFALL, W.; SCHARF, G.; SCHULTE-HERMANN, R. Effects of coffee and its chemopreventive components kahweol and cafestol on cytochrome P450 and sulfotransferase in rat liver. **Food and Chemical Toxicology.**, Oxford, v.46, p.1230-1238, 2008.

IMAISHI, H. MATUMOTO, S. Isolation and functional characterization in yeast of CYP72A18, a rice cytochrome P450 that catalyzes omega-hydroxylation of the herbicide pelargonic acid. **Pesticide, Biochemistry & Physiology**, v.88, p.71-77, 2007.

JAVOT, H.; LAUVERGEAT, V.; SANTONI, V.; MARTIN-LAURENT, F.; GÜÇLÜ, J.; VINH, J.; HEYES, J.; FRANCK, K. I.; SCHÄFFNER, A. R.; BOUCHEZ, D.; MAUREL, C. Role of a single aquaporin isoform in root water uptake. **The Plant Cell**, v.15, p.509-522, 2003.

KAI, K.; TAKAHASHI, H.; SAGA, H.; OGAWA, T.; KANAYA, S.; OHTA, D. Metabolomic characterization of the possible involvement of a Cytochrome P450, CYP81F4, in the biosynthesis of indolic glucosinolate in *Arabidopsis*. **Plant Biotechnology**, v.28, p.379-385, 2011.

KOO, A. J. K.; COOKE, T. F.; HOWE, G. A. Cytochrome P450 CYP94B3 mediates catabolism and inactivation of the plant hormone jasmonoyl-L-isoleucine. **Proceedings of the National Academy of Sciences of the United States of America**, v.108, n.22, p.9298-9303, 2011.

LEE, K. J.; JEONG, H. G. Protective effects of kahweol and cafestol against hydrogen peroxide-induced oxidative stress and DNA damage. **Toxicology Letters**, Shannon, v. 173, n. 2, p. 80-87, 2007.

LI, Z.; HAO, T.; YANG, Y. DENG, W.; Molecular cloning and expression analysis of a cytochrome P450 gene in tomato. **Plant Growth Regulation**, v.61, n.3, p.297-304, 2010.

LIVAK, K. J.; SCHMITTGEN, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the  $2^{-\Delta\Delta CT}$  Method. **Methods**, v.25, p.402-408, 2011.

LUPIEN, S.; KARP, F.; WILDUNG, M.; CROTEAU, R. Regiospecific cytochrome P450 limonene hydroxylases from mint (*Mentha*) species: cDNA isolation, characterization, and functional expression of (-)-4S-limonene-3-hydroxylase and (-)-4S-limonene-6-hydroxylase. **Archives of Biochemistry and Biophysics**, v.368, n.1, p.181-192, 1999.

MORRONE, D.; CHEN, X.; COATES, R. M.; PETERS, R. J. Characterization of the kaurene oxidase CYP701A3, a multifunctional cytochrome P450 from gibberellin biosynthesis. **Biochemical Journal**, v.431, p.337-344, 2010.

NAIDOO, N., CHEN, C.; REBELLO, S.; SPEER, K.; TAI, E. S.; LEE, J.; BUCHMANN, S.; KOLLING-SPEER, I.; VAN DAM, R. M. Cholesterol-raising diterpenes in types of coffee commonly consumed in Singapore, Indonesia and India and associations with blood lipids: A survey and cross sectional study. **Nutrition Journal**, v.48, n. 10, 2011.

OBERTHUER, A.; HERO, B.; SPITZ, R.; BERTHOLD, F.; FISCHER, M. The tumor-associated antigen PRME is universally expressed in high-stage neuroblastoma and associated with poor outcome. **Clinical Cancer Research**, v.10, p.4307-4313, 2004.

PANT, B. D.; BUHTZ, A.; KEHR, J.; SCHEIBLE, W. MicroRNA399 is a long-distance signal for the regulation of plant phosphate homeostasis. **The Plant Journal**, n.53, p.731-738, 2008.

PFALZ, M.; VOGEL, H.; KROYMANN, J. The gene controlling the indole glucosinolate modifier1 quantitative trait locus alters indole glucosinolate structures and aphid resistance in *Arabidopsis*. **The Plant Cell**, v.21, p.985-999, 2009.

PINOT, F.; KANDEL, S.; SAUVEPLANE, V.; COMPAGNON, V.; FRANKE, R.; MILLET, Y.; SCHREIBER, L.; WERCK-REICHHART, D.; Characterization of a methyl jasmonate and wounding-responsive cytochrome P450 of *Arabidopsis thaliana* catalyzing dicarboxylic fatty acid formation *in vitro*. **The Federation of European Biochemical Societies Journal**, v. 274, p. 5116-5127, 2007.

POSTE, S. M.; ROOS, B.; VERMEULEN, M.; AFMAN, L.; JONG, M. C.; DAHLMANS, V. E. Cafestol increases serum cholesterol levels in apolipoprotein E\*3-Leiden transgenic mice by suppression of bile acid synthesis. **Arteriosclerosis, Thrombosis and Vascular Biology**, Dallas, v. 20, n. 6, p.155-156, 2000.

POT, D.; FERREIRA, L. P.; DIAS, R. C. E.; DURAND, N.; GUYOT, B.; RAMOS, J.; SANDRIN-GARCIA, P.; BENASSI, M. T.; MARRACCINI, P.; PEREIRA, L. F. P.; LEROY, T.; VIEIRA, L. G. E. Genetic and molecular determinism of diterpene metabolism in *Coffea* spp. In: **Proceedings of 22<sup>nd</sup> International Conference on Coffee Science (ASIC)**, Campinas, p.174, 2008.

RALSTON, L.; KWON, S. T.; SCHOENBECK, M.; RALSTON, J.; SCHENK, D. J.; COATES, R. M.; CHAPPELL, J. Cloning, heterologous expression, and functional characterization of 5-epi-aristolochene-1,3-dihydroxylase from Tobacco (*Nicotiana tabacum*). **Archives of Biochemistry and Biophysics**, v.393, n.2, p.222-235, 2001.

RAMAKERS, C.; RUIJTER, J. M.; DEPREZ, R. H. L.; MOORMAN, A. F. M.; Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. **Neuroscience Letters**, n.339, p. 62-66, 2003.

ROBERTS, S. C.; Production and engineering of terpenoids in plants cell culture. **Nature Chemical Biology**, v.3, n.7, 2007.

SCHOLZ, M.; LIPINSKI, M.; LEUPOLD, M.; LUFTMANN, H.; HARIG, L.; OFIR, R.; FISHER, R.; PRÜFER, D.; MÜLLER, K. J. Methyl jasmonate induced accumulation of kalopanaxsaponin I in *Nigella sativa*. **Phytochemistry**, v.70, p.517-522, 2009.

SCHRÖDER, G.; UNTERBUSCH, EL.; KALTENBACH, M.; SCHIMIDT, J.; STRACK, D.; DE LUCA, V.; SCHÖDER, J. Light-induced cytochrome P450-dependent enzyme in indole alkaloid biosynthesis: tabersonine 16-hydroxylase. **The Federation of European Biochemical Societies Letters**, v.458, n.2, p.97-102, 1999.

SIMINSZKI, B.; CORBIN, F. T.; WARD, E. R.; FLEISCHMANN, J.; DEWEY, R. Expression of a soybean cytochrome P450 monooxygenase cDNA in yeast and tobacco enhances the metabolism of phenylurea herbicides. **Proceedings of the National Academy of Sciences of the United States of America**, v.96, n.4, p.1750-1755, 1999.

SPEER, K.; SEHAT, N.; MONTAG, A. Fatty acids in coffee. In: **Proceedings of the 15<sup>th</sup> ASIC Colloquium**. Montpellier, p.583-592, 1993.

SPEER, K.; KOLLING-SPEER, I. The lipid fraction of the coffee bean. **Brazilian Journal of Plant Physiology**, Pelotas, v. 18, n. 1, p. 201–216, 2006.

SRIDEVI, V.; GIRIDHAN, P.; RAVISHANKAR, G. A.; Free diterpenes cafestol and kahweol in beans and in vitro cultures of *Coffea* species. **Current Science**, v.99, n.8, p.1001-1004, 2010.

STUMP, M.; FEUSSNER, I. Formation of oxylipins by CYP74 enzymes. **Phytochemistry Reviews**, v.5, p.347-357, 2006.

SWAMINATHAN, S.; MORRONE, S.; WANG, Q.; FULTON, D. B.; PETERS, R. J. CYP76M7 is an ent-cassadiene C11a-hydroxylase defining a second multifunctional diterpenoid biosynthetic gene cluster in Rice. **The Plant Cell**, v.21, p.3315-3325, 2009.

THOLL, D.; SOHRABI, R.; HUH, J.; LEE, S. The biochemistry of homoterpenes – common constituents of floral and herbivore-induced plant volatile bouquets. **Phytochemistry**, v.72, p.1635-1646, 2011.

URGERT, R.; VAN DER WEG, G.; KOSMEIJER-SCHUIL, T. G.; VAN DE BOVENKAMP, P.; HOVENIER, R.; KATAN, M. B. Levels of the cholesterol-elevating diterpenes cafestol and kahweol in various coffee brews. **Journal of Agricultural and Food Chemistry**, Washington, v. 43, n. 8, p. 2167-2172, 1995.

WANG, Q.; HILLWIG, M. L.; WU, Y. PETERS, R. J. CYP701A8: A rice ent-kaurene oxidase paralog diverted to more specialized diterpenoid metabolism. **Plant Physiology**, v.158, n.1, 2012.

WANG, E.; WANG, R.; DEPARASIS, J.; LOUGHRIN, J.H.; GAN, S.; WAGNER, G.J. Suppression of a P450 hydroxylase gene in plant trichome glands enhances natural-product-based aphid resistance. **Nature Biotechnology**, v.19, n.4, p.371–374, 2001.

WILLE, A.; ZIMMERMANN, P.; VRANDÓVA, E.; FÜRHOZ, A.; LAULE, O.; BLEULER, S. HENNIG, L.; PRELIC, A.; ROHT, P. V.; THIELE, L.; ZITZLER, E.; GRUISSEM, W.; BÜHLMANN, P. Sparse graphical gaussian modeling of the isoprenoid gene network in *Arabidopsis thaliana*. **Genome Biology**. V. 5, n. 1, 2004.

XU, W.; BAK, S.; DECKER, A.; PAQUETTE, S. M.; FEYEREISEN, R.; GALBRAITH, D. W. Microarray-based analysis of gene expression in very large gene families: the cytochrome P450 gene superfamily of *Arabidopsis thaliana*. **Gene**, v.272, p.61-74, 2001.

ZHOU, N.; TOOTLE, T. L.; GLAZEBROOK, J. Arabidopsis *PAD3*, a gene required for camalexin biosynthesis, encodes a putative cytochrome P450 monooxygenase. **The Plant Cell**, v.11, p.2419-2428, 1999.

## 6 CONCLUSÃO E PERSPECTIVAS

Este trabalho representa um primeiro esforço para elaborar um catálogo geral da superfamília das Cyt P450s transcricionalmente ativas em cafeeiro, que possivelmente estão envolvidas nas mais variadas vias metabólicas. Os dados servem de base fundamental para estudos bioquímicos e de atividades transcricionais voltados à associação de Cyt P450s com vias metabólicas de interesse agrônômico relacionadas com propriedades nutracêuticas, qualidade da bebida, resistência a estresses bióticos e tolerância a estresses abióticos.

As análises *in silico* identificaram e classificaram 84 contigs de acordo com suas diversas funções associadas as suas subfamílias gênicas, dos quais foram selecionados oito genes candidatos relacionadas à biossíntese de diterpenos, para estudos de perfis transcricionais.

Foi possível também realizar um estudo *in silico* mais aprofundado para os genes da CYP73 (C4H) e CYP98 (C3'H), onde foram analisados a diversidade nucleotídica entre espécies de cafeeiros (*Coffea arabica* e *C. canephora*) e dentro de uma mesma espécie (*C. arabica*), permitindo também para integrar os dados em estudo de mapeamento físico e genético desses genes.

As quantificações por CLAE permitiram identificar tecidos e estágios de desenvolvimento dos frutos de cafeeiros com teores contrastantes de cafeol e cafestol, assim como indicar um perfil de produção de proteína ao longo do desenvolvimento do fruto e sob a ação do MJ.

O perfil transcricional dos oito genes candidatos avaliados apresentou um perfil de expressão bastante particular entre diferentes tecidos ou diferentes fases do desenvolvimento: dois genes tiveram maior atividade transcricional em folha (C22\_1 e C25\_1), três em flor (C19\_2, C28\_1 e C37\_2) e dois em frutos (C33\_1 e C107\_1), e um gene não apresentou expressão específica para nenhum tecido (C23\_1). Os genes C19\_2, C28\_1, C33\_1, C37\_2 e C107\_1 apresentaram um padrão transcricional interessante e possuem grande potencial de estarem envolvidos nas etapas finais da via metabólica de Caf e Cav.

Para tentar elucidar com maior eficiência o padrão de expressão gênica dessa via metabólica, foram selecionados 12 tecidos de cafeeiros para análise de seqüenciamento em larga escala (RNA-Seq), contrastantes para concentração de diterpenos, identificados neste trabalho: folha, flor, perisperma de

frutos (30, 60, 90, 120 e 150DAF) e os frutos controle (0, 24 e 48h) e com aplicação de MJ (24 e 48h).

Com isto, espera-se aumentar o conhecimento sobre os genes relacionados ao acúmulo/degradação de Caf e Cav, e como consequência propor estratégias aplicadas ao melhoramento genético, através da modificação de cafeeiros para estes compostos químicos por biotecnologia, alterando as características nutracêuticas dos frutos para a melhoria da qualidade da bebida.

## REFERÊNCIAS BIBLIOGRÁFICAS

**ABIC**. Associação Brasileira da Indústria de Café. 2010. Disponível em: <<http://www.abic.com.br>>. Acesso em 12 ago 2011.

CAVIN, C.; HOLZÄUSER, D.; CONSTABLE, A.; HUGGETT, A. C.; SCHILTER, B. The coffee-specific diterpenes cafestol and kahweol protect against aflatoxin B1-induced genotoxicity through a dual mechanism. **Carcinogenesis**, Oxford, v. 19, n.8, p. 1369-1375, 1998.

CAVIN, C.; HOLZHAUSER, D.; CONSTABLE, A.; HUBET, AC.; SCHILTER, B. Cafestol and Kahweol, two specific diterpenes with anticarcinogenic activity. **Food and Chemical Toxicology**, Oxford, v. 40, p. 1155-1163, 2002.

**CECAFÉ**. Conselho dos Exportadores de Café do Brasil. 2011. Disponível em: <<http://www.cecafe.com.br>>. Acesso em: 12 ago 2011.

CLARKE, R.J. (1985) Green coffee processing. In: **Coffee: botany, biochemistry and production of beans and beverage**. CLIFFORD, M. N., WILSON, K. C. (Eds), WESPORT (CO). AVI Publishing, p.230-250, 1995.

CROTEAU, R.; KUTCHAN, T. M.; LEWIS, N. G. Natural products (secondary metabolites). In: Buchanan, B.; Gruissem, W.; Jones, R. eds. Biochemistry and Molecular Biology of Plants. **American Society of Plant Biologists**, p. 1250-1268, 2000.

**CONAB**. Companhia Nacional de Abastecimento. 2010. Disponível em: <<http://www.conab.gov.br>>. Acesso em 12 ago 2011.

DE CASTRO, R. D.; MARRACCINI, P. Cytology, biochemistry and molecular changes during coffee fruit development. **Brazilian Journal of Plant Physiology**, v. 18, n.1, 2006.

DELLAPENNA, D.; TIAN, L.; MUSETTI, V.; KIM, J.; MAGALLANES-LUNDBACK, M.; The *Arabidopsis* LUT1 locus encode a member of the cytochrome P450 family that is required for carotenoid  $\epsilon$ -ring hydroxylation activity. **Proceedings of the National Academy of Sciences USA**, v. 101, n. 1, p. 402-407, 2004.

DIAS, R. C. E.; CAMPANHA, F. G.; VIEIRA, L. G. E.; FERREIRA, L. P.; POT, D.; MARRACCINI, P.; BENASSI, M. T.; Evaluation of Kahweol and Cafestol in Coffee Tissues and Roasted Coffee by a New High-Performance Liquid Chromatography Methodology. **Journal of Agricultural and Food Chemistry**, v. 58, n. 1, p. 88-93, 2010.

DUSSERT, S.; CHABRILLANCHE, N.; ROCQUELIN, G.; ENGELMAN, F.; LOPEZ, M.; HAMON, S. Tolerance of coffee (*Coffea* spp.) seeds to ultra-low temperature exposure in relation to calorimetric properties of tissue water, lipid composition and cooling procedure. **Physiologia Plantarum**, n.112, v.4, p. 495-504, 2011.

ERREIRA, L. P.; DIAS, R. C. E.; DURAND, N.; GUYOT, B.; RAMOS, J.; PERTHUIS, B.; SANDRIN-GASRCIA, P.; BENASSI, M. T.; MARRACINI, P.; PEREIRA, L. F. P.; LEROY, T.; VIEIRA, L. G. E.; POT, D. Determinismo genético e molecular do metabolismo de diterpenos em *Coffea* spp. In: Vi Simpósio de Pesquisas dos Cafés do Brasil, Vitória, 2009.

FOLSTAR, P. Lipids. In: CLARKE, R.J. & MACRAE, R. (Eds.) **Coffee: Chemistry**. Londres: Elsevier Applied Science, cap.6, p.203-222, 1985.

GEROMEL, C.; FERREIRA, L. P. F.; GUERREIRO, S. M. C.; CAVALARI, A. A.; POT, D.; PEREIRA, L. F. P.; LEROY, T.; VIEIRA, L. G. E.; MAZZAFERA, P. MARRACINI, P.; Biochemical and genomic analysis of sucrose metabolism during coffee (*Coffea arabica*) fruit development. **Journal of Experimental Botany**, v. 57, n. 12, p.3243-3258, 2006.

GROLLIER, J. F.; PLESSIS, S.; US Patent 4,793, 990, 1998.

HIGGINS, L. G.; CAVIN, C.; ITOH, K.; YAMAMOTO, M.; HAYES, J. D. Induction of cancer chemopreventive enzymes by coffee is mediated by transcription factor Nrf2. Evidence that the coffee-specific diterpenos cafestol and kahweol confer protection against acrolein. **Toxicology and Applied Pharmacology**, San Diego, v. 226, n. 3, p. 328-337, 2008.

HUBER, W. W.; ROSSMANITH, W.; GRUSCH, M.; HASLINGER, E.; PRUSTOMERSKY, S.; PETER-VÖTÖSMARTY, B.; PARZEFALL, W.; SCHARF, G.; SCHULTE-HERMANN, R. Effects of coffee and its chemopreventive components kahweol and cafestol on cytochrome P450 and sulfotransferase in rat liver. **Food and Chemical Toxicology**., Oxford, v.46, p.1230-1238, 2008.

KITZBERGER, C. S. G.; SCHOLZ, M. B. S.; PEREIRA, L. F. P.; VIEIRA, L. G. E.; SERA, T.; SILVA, J. B. G. D.; BENASSI, M. T. Anaysilis of diterpens in green and roasted Coffe of *Coffea arabica* cultivars growing in the same edapho-climatic conditions. In: **Proceedings of 23<sup>rd</sup> International Conference on Coffee Science (ASIC)**, Bali, p. 110, 2010.

KURZOCK, T. SPEER, K. Diterpenes and diterpene ester in Coffee. *Food Reviews International*, v.17, n.4, p.433-450, 2001.

KY, C.L.; LOUARN, J.; DUSSERT, S.; GUYOT, B.; HAMON, S.; NOIROT, M. Caffeine, trigonelline, chlorogenic acids and sucrose diversity in wild *Coffea arabica* L. and *C. canephora* P. accessions. **Food Chemistry**, v.75, p.223-230, 2001.

LANGE, B. M.; GHASSEMIAN, M. Genome organization in *Arabidopsis thaliana*: a survey for genes involved in isoprenoid and chlrophyll metabolism. **Plant Molecular Biology**, v. 51, p. 925 – 948, 2003.

LEE, K. J.; JEONG, H. G. Protective effects of kahweol and cafestol against hydrogen peroxide-induced oxidative stress and DNA damage. **Toxicology Letters**, Shannon, v. 173, n. 2, p. 80-87, 2007.

MASTEN, S.; TICE, R. Cafestol and Kahweol: review of toxicological literature. IN: **Toxicological Summary for Cafestol and Kahweol**, p.44, 1999.

MCGARVEY, D. J.; CROTEAU, R.; Terpenoid metabolism. **Plant Cell**, v. 7, p. 1015-1026, 1995.

NAIDOO, N., CHEN, C.; REBELLO, S.; SPEER, K.; TAI, E. S.; LEE, J.; BUCHMANN, S.; KOLLING-SPEER, I.; VAN DAM, R. M. Cholesterol-raising diterpenes in types of coffee commonly consumed in Singapore, Indonesia and India and associations with blood lipids: A survey and cross sectional study. **Nutrition Journal**, v.48, n. 10, 2011.

NELSON, D. R.; The Cytochrome P450 Homepage. **Human Genomics**, v.4, n.1, p.59-65, 2009. Disponível em: <<http://drnelson.uthsc.edu/cytochromeP450.html>>. Acesso em: 21 ago 2011.

NELSON, D.R. Progress in tracing the evolutionary paths of cytochrome P450. **Biochimica et Biophysica (BBA) – Proteins & Proteomics**, v.1814, n.1, p.14-18, 2011.

NELSON, D. R.; WERCK-REICHHART, D. The Cytochrome P450-centric view of plant evolution. **The Plant Journal**, v.66, p.194-211, 2011.

PINOT, F.; KANDEL, S.; SAUVEPLANE, V.; COMPAGNON, V.; FRANKE, R.; MILLET, Y.; SCHREIBER, L.; WERCK-REICHHART, D.; Characterization of a methyl jasmonate and wounding-responsive cytochrome P450 of *Arabidopsis thaliana* catalyzing dicarboxylic fatty acid formation *in vitro*. **The Federation of European Biochemical Societies Journal**, v. 274, p. 5116-5127, 2007.

POSTE, S. M.; ROOS, B.; VERMEULEN, M.; AFMAN, L.; JONG, M. C.; DAHLMANS, V. E. Cafestol increases serum cholesterol levels in apolipoprotein E\*3-Leiden transgenic mice by suppression of bile acid synthesis. **Arteriosclerosis, Thrombosis and Vascular Biology**, Dallas, v. 20, n. 6, p.155-156, 2000.

POT, D.; FERREIRA, L. P.; DIAS, R. C. E.; DURAND, N.; GUYOT, B.; RAMOS, J.; SANDRIN-GARCIA, P.; BENASSI, M. T.; MARRACCINI, P.; PEREIRA, L. F. P.; LEROY, T.; VIEIRA, L. G. E. Genetic and molecular determinism of diterpene metabolism in *Coffea* spp. In: **Proceedings of 22<sup>nd</sup> International Conference on Coffee Science (ASIC)**, Campinas, p. 174, 2008.

ROBERTS, S. C.; Production and engineering of terpenoids in plants cell culture. **Nature Chemical Biology**, v.3, n.7, 2007.

SCHULER, M.; DUAN, H.; BILGIN, M.; ALI, S. *Arabidopsis* cytochrome P450s through the looking glass: a window on plant biochemistry. **Phytochemistry**, v.5, p.205-237, 2006.

SPEER, K.; SEHAT, N.; MONTAG, A. Fatty acids in coffee. In: **Proceedings of the 15<sup>th</sup> ASIC Colloquium**. Montpellier, p.583-592, 1993.

SPEER, K.; KOLLING-SPEER, I. The lipid fraction of the coffee bean. **Brazilian Journal of Plant Physiology**, v. 18, n. 1, p. 201–216, 2006.

SPILLER, M. A.; The chemical Components off coffee. In: SPILLER, G. A. **Caffeine**. New York: CRC Press, p. 98-161, 1998.

SRIDEVI, V.; GIRIDHAN, P.; RAVISHANKAR, G. A.; Free diterpenes cafestol and kahweol in beans and in vitro cultures of *Coffea* species. **Current Science**, v.99, n.8, p.1001-1004, 2010.

THEIS, N.; LERDAU, M. The evolution of function in plant secondary metabolites. **International Journal of Plant Sciences**, v. 164, n. 3, p. 93-101, 2003.

TISKI, I.; MARRACINI, P.; POT, D.; VIEIRA, L. G. E.; PEREIRA, L. F. P.; Characterization and expression of two cDNA encoding 3-hidroxy-3-methylglutaryl coenzyme A reductase isoforms in Coffee (*Coffea arabica* L.). **Journal of Integrative Biology/OMICS**, Larchmont, v. 15, n.0, p. 110713125909008, 2011.

TUDZYNSKI, B.; HEDDEN, P.; CARRERA, E.; GASKIN, P. The P450-4 Gene of *Gibberella fujikuroi* encodes ent-kaurene oxidase in the gibberellins biosynthesis pathway. **Applied and Environmental Microbiology**, v.67, n.8, p.3514-3522, 2001.

URGERT, R.; VAN DER WEG, G.; KOSMEIJER-SCHUIL, T. G.; VAN DE BOVENKAMP, P.; HOVENIER, R.; KATAN, M. B. Levels of the cholesterol-elevating

diterpenes cafestol and kahweol in various coffee brews. **Journal of Agricultural and Food Chemistry**, Washington, v. 43, n. 8, p. 2167-2172, 1995.

VEZALLI, F.; Agricultura familiar gera empregos mas recebe poucos recursos. 2006. In: Repórter Brasil. Disponível em: <<http://www.reporterbrasil.com.br>>. Acesso em: 3 out 2011.

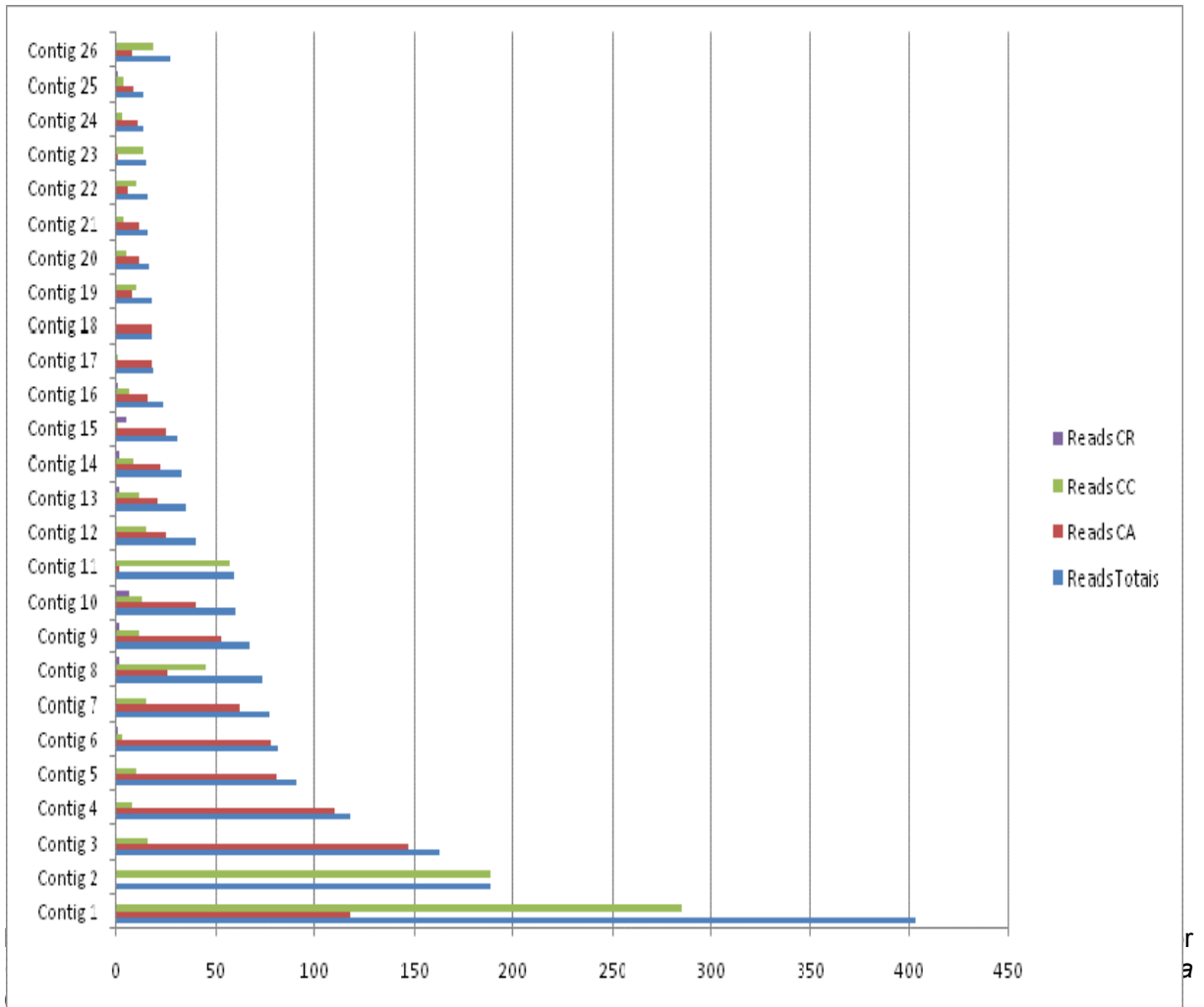
WANG, Q.; HILLWIG, M. L.; WU, Y. PETERS, R. J. CYP701A8: A rice ent-kaurene oxidase paralog diverted to more specialized diterpenoid metabolism. **Plant Physiology**, v.158, n.1, 2012.

WILLE, A.; ZIMMERMANN, P.; VRANDÓVA, E.; FÜRHOZ, A.; LAULE, O.; BLEULER, S. HENNIG, L.; PRELIC, A.; ROHT, P. V.; THIELE, L.; ZITZLER, E.; GRUISSEM, W.; BÜHLMANN, P. Sparse Graphical Gaussian Modeling of the Isoprenoid Gene Network in *Arabidopsis thaliana*. **Genome Biology**. V. 5, n. 1, 2004.

## **ANEXOS**

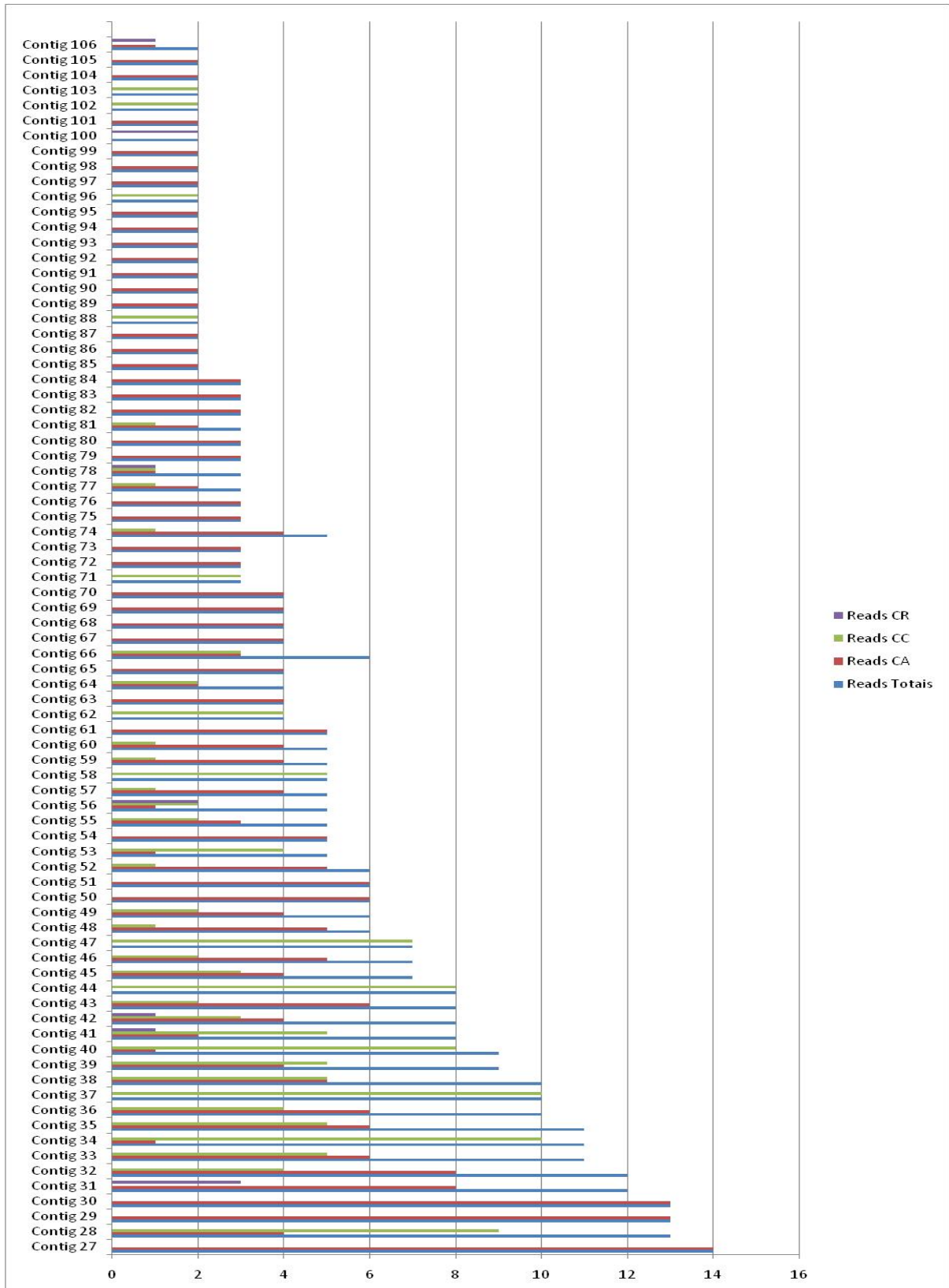
## ANEXO A

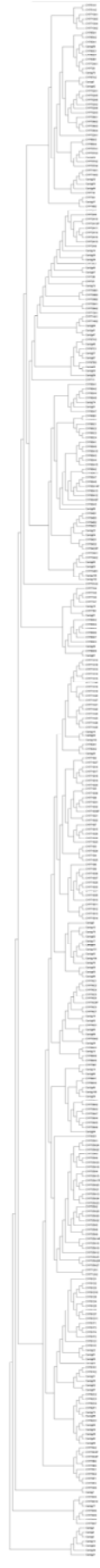
### Contigs formados para Cyt P450 e suas sequências



## ANEXO B

## Contigs formados para Cyt P450 e suas sequências



**ANEXO C - Árvore Filogenética com as Cyt P450 de *Coffea* spp. e *A. thaliana***

## ANEXO D

Artigo Aceito pela revista *Coffee Science* em Janeiro de 2012

### DIVERSIDADE NUCLEOTÍDICA DOS GENES ENVOLVIDOS NA BIOSÍNTESE DE ÁCIDOS CLOROGÊNICOS

**RESUMO:** Os ácidos clorogênicos (CGAs) são compostos químicos importantes de *Coffea* spp. para a qualidade da bebida, pois eles inferem na adstringência e podem alterar o aroma e sabor da bebida. Aproximadamente 310.000 ESTs de *Coffea* estão disponíveis e possibilitam o acesso à variabilidade nucleotídica da planta e o desenvolvimento de marcadores moleculares ligados a qualidade da bebida para as principais enzimas da via de biossíntese dos CGAs: PAL, C4H, 4CL, CQT e C3'H. Neste trabalho foram detectados polimorfismos dos tipos SNP, INDEL ou SSR dentro das sequências nucleotídicas disponíveis no Projeto Genoma Café e no NCBI. As sequências de ESTs de CGAs foram clusterizadas pelo programa Codon Code Aligner, assim como a detecção de polimorfismos e validação dos mesmos (qualidade de cromatograma). Foram identificadas seis isoformas para PAL, uma para C4H, seis para 4CL, duas para CQT e duas para C3'H. Os contigs formados apresentaram um total de 248 polimorfismos (236 SNPs e 12 INDELS), sendo 201 na região codante (127 não sinônimos e 74 sinônimos). A frequência dos polimorfismos foi maior nas regiões UTRs (1pol/54pb) em relação à codante (1pol/81pb). A análise das sequências de *C. arabica* permitiu a identificação de 2 subgrupos diferentes de sequências, referentes aos seus genomas ancestrais (*C. canephora* e *C. eugenoides*). Foi observada a presença de 67,4% dos polimorfismos entre os grupos ancestrais e 32,6% dentro dos grupos em *C. arabica*. Esses resultados vêm permitindo definir genes tanto para estudos de expressão de homeólogos de CGAs como para o desenvolvimento de marcadores moleculares para o mapeamento genético.

Palavras-chave: polimorfismos, marcadores moleculares, SNPs, SSRs, ESTs, CGAs

### NUCLEOTIDE DIVERSITY OF GENES RELATED TO THE CHLOROGENIC ACIDS BIOSYNTHESIS

**ABSTRACT:** The chlorogenic acids (CGAs) are important chemical compounds of *Coffea* spp. related to the quality of the drink. They are related to coffee cup astringency and can change its aroma and flavor. About 310,000 of *Coffea* ESTs are available in public databases providing access to nucleotide variability of the plant and to the development of molecular markers linked to the cup quality of drink for the CGAs enzymes involved in the biosynthesis of CGAs: PAL, C4H, 4CL, CQT and C3'H. There were identified SNPs, INDELSs and SSRs polymorphism related to CGAs from ESTs available from the Brazilian Coffee Genome database and from NCBI. The ESTs sequences for CGAs were trimmed and clustered by the program Codon Code Aligner, as well as the detection of polymorphisms and their validation (quality of chromatogram). We identified six isoforms for

PAL, one for C4H, six for 4CL, two for CQT and two for C3'H. The contigs formed were analyzed for the presence of polymorphisms. A total of 248 polymorphisms (236 SNPs and 12 INDELs) were founded, 201 in the coding region (127 non synonymous and 74 synonymous). The frequency of polymorphisms was higher in the UTR region (1pol/54pb) in relation to coding region (1pol/81pb). The sequences analysis of *C. arabica* allowed the identification of two different subgroups of sequences, related to the ancestrals (*C. canephora* and *C. eugenioides*), the polymorphisms observed between and inside the group were 67,4% and 32,6% respectively. The characterization of the nucleotide diversity on those genes allowed studies on differential expression of homeologous genes as well the use of CGAs as molecular marker related for genetic mapping.

Key-words: polymorfisms, molecular markers, SNPs, SSRs; ESTs, CGAs

## 1 INTRODUÇÃO

O cafeeiro pertence à família Rubiceae e ao gênero *Coffea* da qual se conhecem aproximadamente 103 espécies. Dentre essas espécies duas possuem uma maior importância econômica, a *Coffea arabica* L. e *Coffea canephora* que representam 70% e 30% do mercado total de café, respectivamente (CECAFE, 2011). A espécie *C. arabica* é originária do oeste da África (Etiópia) e é uma planta alotetraplóide ( $2n=4x=44$ ), as demais espécies são diplóides ( $2n=2x=22$ ) e alógamas. A origem provável de *C. arabica* foi uma hibridização de *C. eugenioides* e *C. canephora* (LASHERMES et al., 1999; DAVIS et al., 2007).

O Brasil é o maior produtor e exportador mundial de café, sendo responsável por 30% do mercado internacional (CECAFE, 2011). A qualidade da bebida de café é o principal fator de agregação de valor ao produto, pois confere melhores preços no mercado e, portanto maior competitividade. A composição química do café é um dos fatores que determinam a qualidade da bebida. Seu sabor e aroma são resultantes da presença combinada de vários constituintes químicos voláteis e não voláteis, entre eles os açúcares, cafeína, trigonelina, lipídeos e os CGAs (FARAH et al., 2006). Estes últimos são responsáveis pela adstringência e interferem no seu sabor (KIM et al., 2009).

Os CGAs são compostos de metabolismos secundários e derivam da biossíntese da fenilalanina pela via dos fenilpropanóides (CAMPA et al., 2004). Existem duas vias de síntese dos CGAs, primeiro a enzima fenilalanina amonialiase (PAL) catalisa a fenilalanina em ácido cinâmico, a cinamato 4 hidroxilase (C4H) catalisa a segunda hidroxilação do ácido cinâmico para ácido coumárico, este é catalizado pela enzima 4 couramato CoA ligase (4CL) em p-coumaril CoA. A partir deste ponto podem ser feitas duas rotas diferentes. Na primeira o ácido p-coumárico pode ser catalizado em ácido cafeico pela 4-couramato 3-hidroxilase

(C3`H) que em seguida é catalisado pela enzima 4CL em Cafeoil CoA, no qual é adicionado a molécula de ácido quínico pela enzima hidroxicinamoil transferase quinato (CQT), originando as moléculas de CGAs. Na segunda, o ácido p-coumárico é adicionado a molécula de ácido quínico pela enzima CQT que origina o ácido p-coumaroil quínico, o qual é catalisado pela enzima C3`H em CGAs (GOPINADHAN, 2008) (Figura 1).

O estudo das enzimas envolvidas nesta via de biossíntese é importante, pois os CGAs são precursores da síntese de lignina que está envolvida na proteção das células da planta contra estresses bióticos e abióticos (resistência a patógenos e redução da permeabilidade da parede celular em relação à água), além de serem fundamentais para a estruturação rígida das plantas (KOCHKO et al., 2003; McCARTHY, 2007). O interesse do uso de CGAs na saúde humana está aumentando devido as suas propriedades antioxidantes, efeito antagonista a opióides e a sua capacidade de transporte de glicose no combate ao diabetes (BAKURADZE, et al., 2011).

Com o desenvolvimento da biologia molecular, surgiram novas técnicas para auxiliar o melhoramento genético, este é o caso dos marcadores moleculares que podem indicar com precisão a variabilidade genética dos indivíduos (BORÉM, 2009). Estes possuem inúmeras vantagens em relação aos marcadores morfológicos, pois independem das condições ambientais e estágio fisiológico da planta, além de permitirem uma identificação precoce de genótipos com características de interesse (EVANS & CARDON, 2004). O mapeamento de genes relacionados aos CGAs tem sido realizado em várias espécies de plantas como a maçã (CHAGNE et al, 2012) e alcachofra (MENIN et al., 2010).

Dentre os vários tipos de marcadores moleculares disponíveis, os mais interessantes para o estudo de mapeamento genético são os microssatélites (SSRs) e os polimorfismos de base única (SNPs). Estes possuem uma maior frequência dentro do genoma dos organismos e permitem a identificação e mapeamento de genes que controlam características de interesse agrônômico, fator interessante para reduzir o tempo e o custo do melhoramento clássico de plantas (BORÉM, 2009).

A utilização de dados de projetos de transcriptoma de café, como ESTs (seqüências de DNA expressas), vem permitindo estudos visando entender melhor o funcionamento genético da planta (MONDEGO et al., 2011; VIDAL et al., 2010). No *Genbank* (NCBI) além das mais de 260.000 seqüências do Genoma Café, estão disponíveis aproximadamente 47 mil seqüências ESTs de *C. canephora* (LIN et al., 2005) e 10 mil provenientes do IRD (Institute de Recherche pour le Developpement) (PONCET et al., 2006).

Este trabalho visa a busca e seleção *in silico* de seqüências das enzimas envolvidas na biossíntese dos CGAs, com o objetivo de identificar o número de isoformas para cada enzima, avaliar o nível de polimorfismos disponível dentro dos ESTs, analisar a origem destes polimorfismos (inter específico ou intra específico), calcular a frequência dos mesmos de acordo com as regiões (UTRs ou codantes) e iniciar um trabalho de mapeamento das isoformas identificadas dentro de populações de *Coffea* spp baseado em análises prévias utilizando ferramentas de bioinformática.

## 2 MATERIAIS E MÉTODOS

As buscas das seqüências de interesse foram feitas na plataforma do Genoma Café (<http://www.lge.ibi.unicamp.br/cafe>) através das ferramentas *BlastX* e no *Genbank/NCBI* (<http://www.ncbi.nlm.nih.gov/>) por *tBlastN* com a utilização de seqüências proteicas codificantes para as enzimas envolvidas na biossíntese de CGAs previamente identificadas no genoma de *Arabidopsis thaliana* ([www.arabidopsis.org](http://www.arabidopsis.org)). Após comparação, foram selecionadas somente seqüências com *e-value* menor que  $e^{-10}$ . As seqüências encontradas foram inseridas no Programa *Codon Code Aligner* (versão 1.6.3) para análise e formação dos *contigs*. O mesmo programa fez a limpeza das seqüências e análise da qualidade dos cromatogramas através das ferramentas denominadas *call base*, *clip ends*, *trim vector* e *find heterozygous* (Phred e Phrap) para excluir regiões de baixa qualidade e descartar as seqüências que correspondem a vetores. Os parâmetros para o alinhamento das seqüências foram a presença de percentual mínimo de identidade de 90% e homologia mínima de 20 pares de base. Após a formação dos *contigs*, a seqüência consenso dos mesmos foi analisada através da ferramenta *BlastX* no *Genbank/NCBI* para validar a codificação da proteína de interesse. Após confirmação, o quadro de leitura da proteína foi determinado pela análise realizada na plataforma online *EXPASY Translate Tools* (<http://www.expasy.ch/tools/dna.html>).

A identificação do número de isoformas para cada enzima foi definido através da comparação das seqüências protéicas dos seus respectivos *contigs* com o uso da ferramenta *bl2seq* (*blast two sequence/Genbank/NCBI*), onde foram consideradas isoformas diferentes quando a porcentagem de homologia e identidade apresentavam valores menores que 90%.

A detecção dos polimorfismos foi feita baseada na qualidade dos cromatogramas e somente quando eram detectados em pelo menos duas seqüências (mínimo de quatro

seqüências por contig). Foram desenhados oligonucleotídeos específicos para analisar preferencialmente as regiões com maior frequência polimórfica observada *in silico*.

### 3 RESULTADOS E DISCUSSÃO

A busca *in silico* de polimorfismos para as enzimas PAL, C4H, 4CL, CQT e C3'H e formação das suas isoformas, selecionou um total de 426 seqüências ESTs do banco de dados do Projeto Genoma Café. O número total de ESTs para cada uma das cinco enzimas de CGAs, assim como a quantidade de contigs formados e número de isoformas encontradas estão descritas na tabela 1.

Tabela 1: Seleção de ESTs e caracterização dos contigs e *singletons* encontrados *in silico*

ENZIMA	ESTs	Contigs*	Singletons	Isoformas	Proteína	
					Completa (Full Length)	Incompleta (Partial Length)
PAL	133	4	2	5	2	2
C4H	89	1	0	1	1	0
4CL	114	6	0	6	3	3
CQT	47	3	0	2	2	1
C3'H	46	2	0	2	2	0

\* Contigs formados com no mínimo 4 seqüências de ESTs

Análises de BlastX dos 16 contigs formadas no *Genbank* (NCBI) resultaram em similaridade com algumas seqüências previamente anotadas para codificação das proteínas relacionadas ao metabolismo de CGAs (Tabela 2). A validação com os dados do *Genbank* é importante visando confirmar a montagem e tamanho dos contigs, assim com sua classificação dentro dos CGAs. Todas as isoformas encontradas neste trabalho foram depositadas no *Genbank* (NCBI). É interessante observar que o número de ESTs das duas primeiras enzimas da via de fenilpropanóides, PAL, C4H e 4CL, foi três vezes maior que o das duas ultimas enzimas envolvidas na formação dos CGAs. Essas enzimas respondem não somente para produção de CGAs, mas também para a produção de uma série de outros compostos como por exemplo flavonóides, ácido salicílico e estilbenos

Tabela 2: Resultado do BlastX dos 16 contigs comparados contra o GenBank (NCBI)

Enzyme	Length of Contigs (bp)	Organism	Accession Number	E-value
--------	---------------------------	----------	------------------	---------

PAL_C1	2673	<i>C. canephora</i>	AAN32866.1	0.0
PAL_C2	1056	<i>C. arabica</i>	AEL21617.1	1 e-130
PAL_C3	2253	<i>C. canephora</i>	AEO94540.1	0.0
PAL_C4	2489	<i>C. canephora</i>	AEO94541.1	0.0
C4H_C4	3349	<i>C. roseus</i>	CAA83552	0.0
4CL_C2	2389	<i>C. arábica</i>	CAJ41420.1	0.0
4CL_C3	2065	<i>R. idaeus</i>	AAF91309.1	0.0
4CL_C4	2109	<i>N. tabacum</i>	AAB18637.1	0.0
4CL_C5	1543	<i>A. thaliana</i>	AAP03021.1	0.0
4CL_C6	1083	<i>N. sylvestris</i>	AAO25512.1	1 e-137
4CL_C7	1307	<i>P. trichocarpa</i>	EEE96927.1	3 e-172
CQT_C1	1097	<i>C. canephora</i>	ABO77957.1	2 e-62
CQT_C4	2169	<i>C. arabica</i>	CAT00081.1	0.0
CQT_C14	1650	<i>C. canephora</i>	ABO47805.1	0.0
C3`H_C25	1939	<i>C. canephora</i>	ABO77958.1	0.0
C3`H_C34	1698	<i>C. canephora</i>	ABO83677	0.0

Os números totais de seqüências ESTs selecionadas pelas buscas *in silico* para cada contig estão representados na Tabela 3. Essas seqüências foram separadas de acordo com as suas respectivas espécies de origem pelo fato de influenciarem as porcentagens de frequência dos polimorfismos.

Tabela 3: Tabela contendo o número total de seqüências e distribuição nas espécies de *Coffea*.

Enzyme	S <sup>a</sup> Total	S <sup>a</sup> (Cc)	S <sup>a</sup> (Ca)	S <sup>a</sup> (Cr)
PAL_C1	34	6	27	1
PAL_C2	7	4	3	0
PAL_C3	52	27	25	0
PAL_C4	38	6	31	1
C4H_C4	89	21	67	1
4CL_C2	60	18	41	1
4CL_C3	19	6	13	0
4CL_C4	20	5	15	0
4CL_C5	6	1	5	0
4CL_C6	2	0	2	0
4CL_C7	7	2	4	0
CQT_C1	2	0	2	0
CQT_C4	37	12	25	0
CQT_C14	8	4	4	0
C3`H_C25	42	17	25	0
C3`H_C34	4	0	4	0

S<sup>a</sup> = Número Total de Seqüências; Cc = *Coffea canephora*; Ca = *Coffea arabica*; Cr = *Coffea racemosa*.

De acordo com os dados apresentados na tabela 3, é possível perceber que em quase todos os contigs há uma predominância de seqüências de *C. arabica*, com exceção do Contig 3 da PAL que apresentou um equilíbrio, indicando que essa isoforma deve possuir uma maior

expressão em *C. canephora*. Este resultado também é devido a presença majoritária de bibliotecas de *C. arabica* no Projeto Genoma Café (VIEIRA et al., 2006).

Nos 16 contigs foram identificados um total de 248 polimorfismos, sendo 47 dentro das regiões UTRs e 201 dentro das regiões codantes. Inseridos nesta última 63% dos polimorfismos (127) correspondem ao tipo sinônimo (S) e 37% deles (74) ao tipo não sinônimo (NS) (Tabela 4). Uma porcentagem semelhante foi verificada para trigo por RAVEL et al, (2006) que encontrou aproximadamente 33% de S e 67% de NS, demonstrando que a maioria das mutações mais comuns não alteram o tipo do aminoácido formado.

Tabela 4: Quantidade e Tipos de Polimorfismos Encontrados.

Enzima	Número de Polimorfismos Totais	SNPs	Indels	Região 5'UTR	Região Codificadora		Região 3'UTR
					Sinônimos	Não Sinônimos	
PAL_C1	23	23	0	0	18	5	0
PAL_C2	2	2	0	1	1	0	0
PAL_C3	43	42	1	0	22	17	4
PAL_C4	32	32	0	0	20	8	4
C4H_C4	34	30	4	0	17	11	6
4CL_C2	22	20	2	0	6	5	11
4CL_C3	4	3	1	0	2	1	1
4CL_C4	25	25	0	3	15	7	0
4CL_C6	0	0	0	0	0	0	0
4CL_C7	0	0	0	0	0	0	0
CQT_C4	24	24	0	1	15	6	2
CQT_C14	4	4	0	0	2	2	0
C3'H_C25	35	31	4	13	9	12	1

Os polimorfismos foram encontrados tanto em regiões não codificadoras de proteínas (5'UTR e 3'UTR), como em regiões codificadora. Os resultados evidenciam a maior frequência de SNPs (95%) em relação aos indels (5%), assim como a de polimorfismos sinônimos (63%) em relação aos não sinônimos (37%), como foi observado por SCHMID et al. (2003). A partir da detecção dos polimorfismos calculou-se as frequências da presença de polimorfismos nas 3 diferentes regiões das seqüências (5'UTR, região codificadora da proteína e 3'UTR) a cada 100 pares de bases, de acordo com o seu respectivo contig. Foi encontrado a frequência de 1 polimorfismo a cada 50 pares de base nas regiões UTRs e 1 polimorfismo a cada 81 pares de base nas regiões codificadoras de proteína.

A pesquisa realizada *in silico* para conhecer a frequência dos polimorfismos (Tabela 5) de acordo com a sua localização é importante, pois pode-se inferir a região que possui a maior probabilidade de encontro dos mesmos nas análises *in vivo*. Com essa informação, é possível realizar o desenho dos oligonucleotídeos específicos priorizando as regiões que apresentem uma frequência maior de polimorfismos para a detecção dos mesmos. Segundo LAI et al., (2008) em análises de ESTs de girassóis, a porcentagem de sucesso de encontrar um polimorfismo *in vivo* foi 72% quando análises prévias *in silico* indicaram as melhores regiões para desenhos de oligonucleotídeos, contra 37% de sucesso quando eles eram desenhados ao acaso.

Os dados de frequência dos polimorfismos evidenciaram que na maioria das vezes a esta se encontra nas regiões não codificadoras de proteína (3'UTR e 5'UTR), assim como foi relatado por MONDEGO et al. (2011). A grande frequência de SNPs encontrados em todas as enzimas estudadas confirma os dados relatados por LIJAVETZKY et al. (2007) que diz que os SNPs são os polimorfismos mais frequentes no genoma de uva. A média de SNPs totais encontrados foi de 1.4 polimorfismos a cada 100 pares de base uma quantidade relativamente alta se comparado aos resultados obtidos por VIDAL et al. (2010) que observou 0.39 polimorfismos, porém isto provavelmente deve-se ao fato de que a região 5'UTR da C3'H\_C25 elevou a média observada. Outro estudo que comprova a maior frequência dos SNPs em relação aos microssatélites (SSRs) foi o realizado por PONCET et al. (2006) que encontrou um SSR a cada 7730pb em ESTs de *C. canephora*, valor quase 5.5 vezes menor do que os SNPs detectados por este trabalho.

Tabela 5: Frequência dos Polimorfismos segundo suas localizações.

Gene	Frequência 5'UTR/100 pb	Frequência Coding/100 pb	Frequência 3'UTR/100 pb	Frequência Média
PAL_C1	NA	1,07	0	1pol/99 bp
PAL_C2	0,50	0,25	NA	1pol/292 bp
PAL_C3	NA	2,22	1,65	1pol/46 bp
PAL_C4	0	1,35	3,30	1pol/78 bp
C4H_C4	NA	1,80	3,52	1pol/54 bp
4CL_C2	NA	0,69	2,82	1pol/93 bp
4CL_C3	NA	0,22	0,86	1pol/429 bp
4CL_C4	3,16	1,35	0	1pol/71bp
4CL_C6	0	NA	NA	0
4CL_C7	NA	0	0	0
CQT_C4	1,03	1,62	0,60	1pol/78 bp
CQT_C14	NA	0,41	0	1pol/52 bp
C3'H_C25	12,38	1,37	0,51	1pol/52 bp

NA= regiões com menos de quatro seqüências.

De acordo com os resultados obtidos na detecção de polimorfismos e identificação dos mesmos, pode-se inferir a existência de diferenças alélicas inter-espécies (*C. arabica* x *C. canephora*, *C. arabica* x *C. racemosa* e *C. canephora* x *C. racemosa*) e intra-espécies (*C. arabica* x *C. arabica*, *C. canephora* x *C. canephora* e *C. racemosa* x *C. racemosa*). Os tipos de polimorfismos SNPs detectados na análise *in silico* para as cinco enzimas estão contidas na Tabelas 6.

Tabela 6: Tipos de Polimorfismos Encontrados *in silico* para as enzimas dos CGAs.

Gene	Contig	Intra Ca	Intra Cc	Intra Cr	Inter Ca e Cc	Inter Ca≠Cr	Inter Cc≠Cr
PAL	1	17	5	NA*	2	0	2
PAL	2	1	2	NA	0	NA	NA
PAL	3	32	20	NA	0	NA	NA
PAL	4	28	0	NA*	0	4	6
C4H	1	26	4	NA*	4	0	7
4CL	1	18	8	NA*	1	0	1
4CL	2	0	1	NA	3	NA	NA
4CL	3	23	3	NA	0	NA	NA
4CL	4	0	0	NA	0	NA	NA
4CL	6	0	0	NA	0	NA	NA
CQT	2	22	2	NA	0	NA	NA
CQT	3	3	0	NA	1	NA	NA
C3`H	1	31	7	NA	0	NA	NA
TOTAL		201	52	0	11	4	16

Ca = *C. arabica*; Cc = *C. canephora*; Cr = *C. racemosa*; NA = não havia seqüências de *C. racemosa*. NA\* = Havia apenas 1 seqüência de *C. racemosa*.

A análise dos polimorfismos inter (E) e intra grupos (I) resultou no encontro de 71% de polimorfismos dentro de *C. arabica*, 18% dentro de *C. canephora*, 3.9% entre *C. arabica* e *C. canephora*, 1.4% entre *C. arabica* e *C. eugenioides* e 5.6% entre *C. canephora* e *C. eugenioides*. Estes dados corroboram com os encontrados por VIDAL et al. (2010), cujo estudo encontrou 81% dos polimorfismos dentro de *C. arabica*, 17% dentro de *C. canephora*, 4% entre *C. arabica* e *C. canephora*, 2.5% entre *C. arabica* e *C. eugenioides*.

Também foram classificados os polimorfismos entre as espécies (31 polimorfismos) e dentro delas (253 polimorfismos). Dentro das espécies foi observado um maior número de polimorfismos em *C. arabica* (71%), seguido de *C. canephora* (18,3%). A maioria dos SNPs encontrados localizaram-se dentro da espécie *C. arabica*, isto devido ao maior número de seqüências desta espécie dentro da maioria dos contigs.

Dentro da espécie de *C. arabica* foi possível separar as seqüências em 2 grupos genômicos diferentes (subgenomas), um deles assemelha-se grupo de *C. canephora* e o outro provavelmente seja semelhante a *C. eugenioides*, isto se deve possivelmente ao fato de que *C.*

*arabica* é um híbrido originado entre o cruzamento de *C. canephora* e *C. eugenoides* (Figura 2). A formação de 2 grupos distintos de seqüências dentro de *C. arabica*, reforçam os trabalhos sobre a origem alopoliplóide da espécie a partir da hibridização das duas espécies diplóides *C. canephora* e *C. eugenoides* (LASHERMES et al., 1999, DAVIS et al., 2007).

Analisando apenas os polimorfismos dentro dos subgenomas de *C. arabica*, foi possível concluir que a maioria deles situava-se entre os dois grupos (136 polimorfismos) e a minoria dentro de um dos dois grupos formados (68 polimorfismos). Essa separação por SNPs é importante, pois permite estudos de caracterização da expressão de genes homeólogos em *Coffea arabica*, identificando a expressão diferencial de genes de um subgenoma com relação ao outro (VIDAL et al., 2010; MARRACCINI et al., 2011). Entretanto visando a identificação de genótipos diferentes e estudos de mapeamento, somente os SNPs intra-específicos podem ser utilizados.

Os resultados gerados com a utilização de ESTs disponíveis nos bancos de dados públicos é um trabalho de base que auxilia e aumenta as probabilidades de se desenvolver marcadores moleculares efetivos para características de interesse agrônômico, pois possibilita inferir a diversidade nucleotídica esperada *in vivo* de indivíduos de uma população, acelerando o processo de busca de marcadores polimórficos. Este fator é de fundamental importância para futuros trabalhos de mapeamento e de seleção assistida por marcadores espécies perenes economicamente importantes, diminuindo custos e tempo de melhoramento clássico da planta (COGAN et al., 2007).

Baseado nessas análises *in silico* algumas populações de *Coffea* spp. estão sendo genotipadas com os *primers* desenvolvidos por este trabalho. Sendo que um deles já apresentou resultados positivos e foi mapeado em *Coffea canephora* (LEROY et al., 2011). Para a população de mapeamento de *Coffea arabica* as análises estão em desenvolvimento.

#### 4 CONCLUSÃO

As análises *in silico* são eficientes para identificar genes da principais via metabólicas dos CGAs de cafeeiro. Foram identificados 16 contigs relacionados a cinco dos principais genes da biossíntese de CGAs. Esses genes apresentam uma alta freqüência de polimorfismos do tipo SNPs (95%), com sua maior distribuição nas regiões UTRs (1pol/50pb) em relação as regiões codificadoras de proteína (1pol/81pb). A análise *in silico* também identificou a presença de dois grupos distintos de seqüências dentro do grupo de *C. arabica*, de acordo com os seus genomas ancestrais, permitindo realizar estudos de expressão de genes homeólogos.

Os genes e polimorfismos encontrados vem sendo utilizados em trabalhos de genotipagem de populações de *Coffea* spp. para mapeamento físico e genético dos CGAs.

## 5 AGRADECIMENTOS

Ao Consórcio Pesquisa Café e FINEP/GENOCAFÉ pelo apoio financeiro. S.T.I. foi bolsista de Mestrado CNPq; L.F.P.P. e L.G.E.V. são bolsistas de produtividade – CNPq.

## 6 REFERÊNCIAS BIBLIOGRÁFICAS

BAKURADZE, T.; BOEHM, N.; JANZOWSKI, N.; LANG, R.; HOFMANN, T.; STOCKIS, J. P.; ALBERT, F. W.; STIEBITZ, H.; BYTOF, G.; LANTZ, I.; BAUM, M.; EISENBRAND, G. Antioxidant-rich coffee reduces DNA damage, elevates glutathione status and contributes to weight control: results from intervention study. **Molecular Nutrition and Food Research**, v.55, p.793-797, 2011.

BORÉM, A. A aplicação dos marcadores moleculares no melhoramento. In: Marcadores Moleculares. Ed. Borém e Caixeta, 2ed., p.95-102, Viçosa: editora UFV, 2009.

CAMPA, C. et al. **Candidate gene strategy for the study of the chlorogenic acid biosynthesis**. Montpellier: 2004

CECAFE. **Conselho dos Exportadores de Café do Brasil**. 2011. Disponível em: <<http://www.cecafe.com.br>>. Acesso em: 12 ago 2011

CHAGNE, D.; KRIEGER, C.; RASSAM, M.; SULLIVAN, M.; FRASER, J.; ANDRE, C.; PINDO, M.; TROGGIO, M.; GARDINER, S. E.; HENRY, R. A.; ALLAN, A. C.; McGHIE, T. K.; LAING, W. A. QTL candidate gene mapping for polyphenolic composition in Apple fruit. **BMC Plant Biology**, v.12, n.12, 2012.

COGAN, N. O.; DRAYTON, M. C.; PONTING, R. C.; VECCHIES, A.C.; BANNAN, N. R.; SAWBRIDGE, T. I.; SMITH, K. F.; SPANGENBERG, G.C.; FORSTER, J. W. Validation of in silico-predicted genic SNPs in white clover (*Trifolium repens* L.), an outbreeding allopolyploid species. **Molecular Genetics Genomics**, v. 277, n. 4, p.89-113, 2007.

DAVIS, A. P.; CHESTER, M.; MAURIN, O.; FAY, M. F. Searching for the relatives of *Coffea* (Rubiaceae, ixoroideae): the circumscription and phylogeny of Coffeae based on plastid sequence data and morphology. **American Journal of Botany**, v.94, n.3, p.313-329, 2007.

EVANS, D. M.; CARDON, L. R.; Guidelines for genotyping in genomewide linkage studies: single-nucleotide-polymorphism maps versus microsatellite maps. **Genetics**, v. 75, p. 687-692, 2004.

FARAH, A.; MONTEIRO, M. C.; CALADO, V.; FRANCA, A. S.; TRUGO, L. C.; Correlation between cup quality and chemical attributes of Brazilian coffee. **Food Chemistry**, Oxford, v. 98, n. 2, p.373-380, 2006.

KIM, J. G.; BEPPU, K.; KATAOKA, I. Varietal differences in phenolic content and astringency in skin and flesh of hardy kiwifruit resources in Japan. **Scientia Horticulturae**, v.120, n. 4, p.551-554, 2009.

KOCHKO, A. De. et al. Genetic mapping of caffeoyl-coenzyme A 3-O-methyltransferase gene in coffee trees: impact in chlorogenic acid content. **Theoretical and Applied Genetics**, p. 751-756, 2003.

LAI, Z.; LIVINGSTONE, K.; ZOU, Y. CHURCH, S. A.; KNAPP, S. J.; ANDREWS, J.; RIESEBERG, L. H. Identification and mapping of SNPs from EST sunflower. **Theoretical and Applied Genetics**, v.111, n.8, p.1532-1544, 2008.

LASHERMES, P.; COMBES, M. C.; ROBERT, J.; TROUSLOT, P.; D'HONT, A.; ANTHONY, F.; CHARRIER, A. Molecular characterization and origin of the *Coffea arabica* L. genome. **Molecular Genome and Genetics**, 261, p. 259-266, 1999.

LEROY, T.; DE BELLIS, F.; LEGNATE, H.; KANAMURA, E.; GONZALES, G.; PEREIRA, L. F.; ANDRADE, A. C.; CHARMETANT, P.; MONTAGNON, C.; CUBRY, P.; MARRACINI, P.; POT, D.; KOCHKO, A. Improving the quality of African robustas: QTLs for yield-and quality-related traits in *Coffea canephora*. **Tree Genetics & Genomes**, v. 7, n. 4, p. 781-798, 2011.

LIJAVETZKY, D.; CABEZAS, J. A.; IBÁÑEZ, A.; RODRÍGUEZ, V.; MATÍNEZ-ZAPATER, J. M. High throughput SNP Discovery and genotyping in grapevine (*Vitis vinifera* L.) by combining a re-sequencing approach and SNPlex technology. *BMC Genomics*, n.8 , v. p.424, 2007.

LIN, C. et al. Coffee and Tomato share common gene repertoires as revealed by deep sequencing of seeds and cherry transcripts. *Theoretical and Applied Genetics*, set. 2005.

MARRACCINI, P. et al. RBCS1 expression in coffee: *Coffea* orthologs, *Coffea arabica* homeologs, and expression variability between genotypes and under drought stress. *BMC Plant Biology*, v.11, art.85, maio 2011.

McCARTHY, J. et al. Chlorogenic acid synthesis in coffee: An analysis of CGA content and real-time RT-PCR expression of HCT, HQT, C3H1, and CCoAOMT1 genes during grain development in *C. canephora*. *Plant Science*, v. 172, p. 861-1060, fev 2007.

MENIN, B.; COMINO, C.; MOGLIA, A.; DOLZHENKO, Y.; PORTIS, E.; LANTERI, S. Identification and Mapping of genes related to caffeoylquinic acid synthesis in *Cynara carduntus* L. *Plant Science*, v.179, p.338-347, 2010.

MONDEGO, J.; VIDAL, R.; CARAZZOLLE, M. F.; Tokuda, E. K.; Parizzi, L. P.; Costa, G. G. L.; Pereira, L. F. P.; ANDRADE, A. C.; Colombo, C.; VIEIRA, L. G. E.; PEREIRA, G. A. G.; Brazilian coffee genome project consortium. An EST-based analysis identifies new genes and reveals distinctive gene expression features of *Coffea arabica* and *Coffea canephora*. *BMC Plant Biology*, v. 11, n. 30, 2011.

PONCET, V. et al. SSR mining in coffee tree EST databases – Potential use of EST-SSR as markers for the *Coffea* genus. *Molecular Genetic and Genomics*, v. 276, n. 5, p. 436-449, nov. 2006.

RAVEL, C.; PRAUD, S.; MURIGNEUX, A.; CANAGUIER, A.; SAPET, F.; SAMSON, D.; BALFOURIER, F.; DUFOUR, P.; CHALHOUB, B.; BRUNEL, D.; BECKERT, M.;

CHARMET, G. Single-nucleotide polymorphism frequency in a set of selected lines of bread wheat (*Triticum aestivum* L.). **Genome**, v.49, p.1131-1139, 2006.

SCHMID, K. J. et al. Large-Scale Identification and Analysis of Genome-Wide Single-Nucleotide Polymorphisms for Mapping in *Arabidopsis thaliana*. **Genome Research**, p. 1250-1257, 2003.

RUPASINGUE, H. P. V. The role of polyphenols in quality postharvesting handling, and processing of fruits. In: Postharvesting biology and technology of fruits, vegetables, and flowers. Ed. GOPINADHAN, P.; MURR, D. P.; HANDA, A. K. LURIE, S. Iowa: Wiley-Blackwell, 2008. p.482.

VIDAL, R.; MONDEGO, J.; POT, D.; AMBROSIO, A. B. ; ANDRADE, A. C.; PEREIRA, L. F. P.; VIEIRA, L. G. E.; CARAZZOLLE, M. F.; PEREIRA, G. A. G. A high-throughput data mining of SNPs in *Coffea* spp ESTs suggests differential homeologous gene expression in the allotetraploid *Coffea arabica*. **Plant Physiology**, Bethesda, v. 154, 2010.

VIEIRA, L. G. E. et al. Brazilian Coffee genome Project: an EST-based genomic resource. **Brazilian Journal of Plant Physiology**, v. 18, p. 95-108, jan/mar 2006.

7 FIGURAS

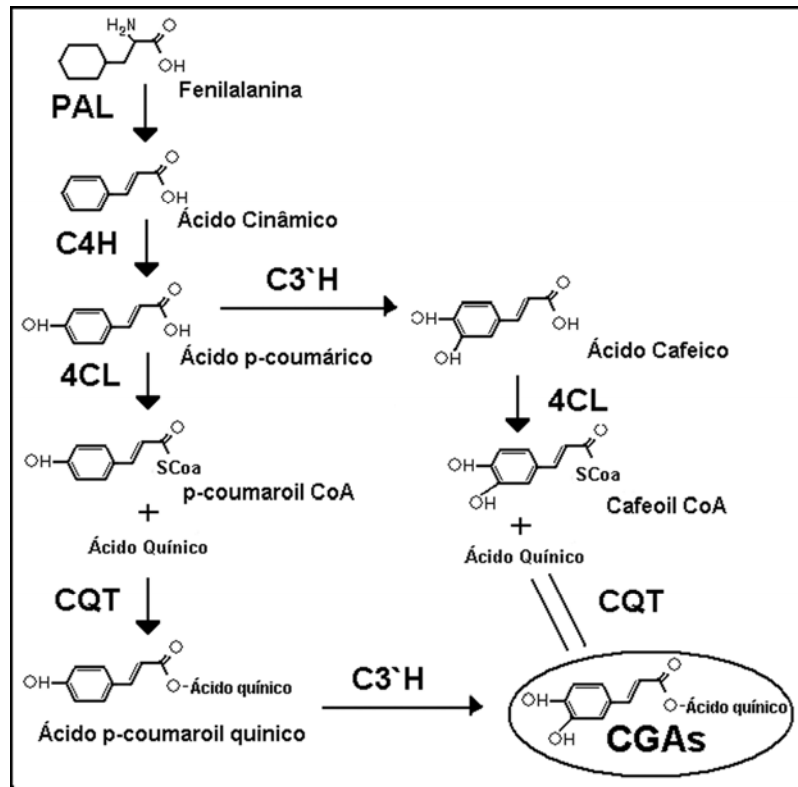
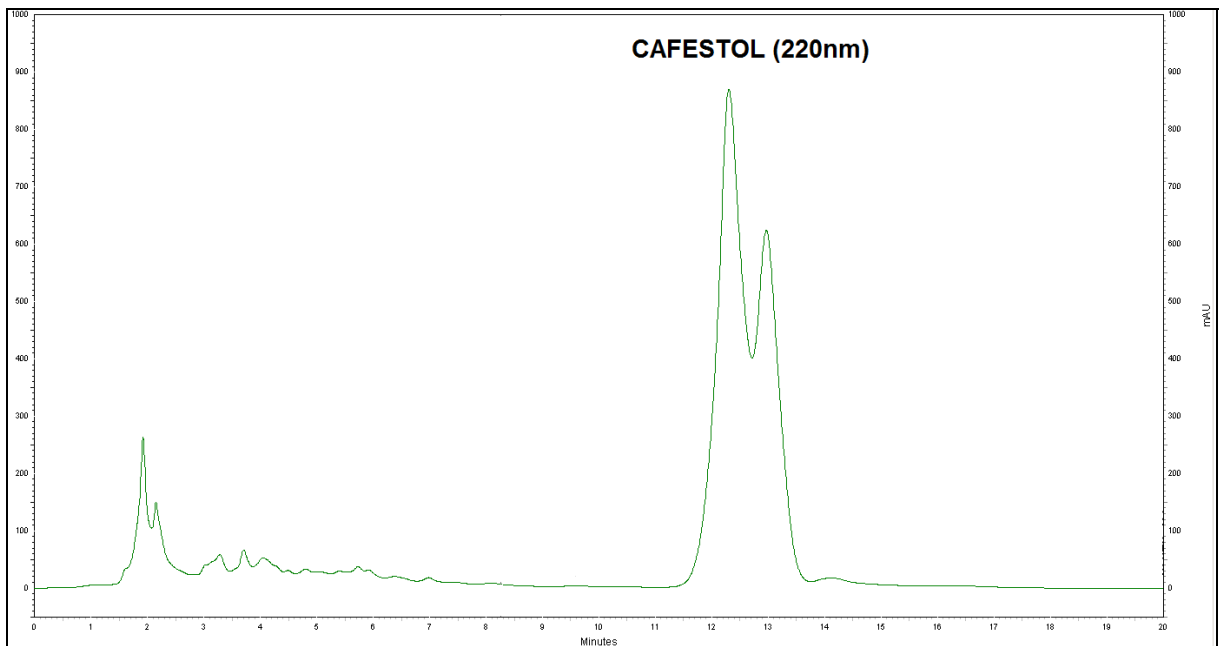


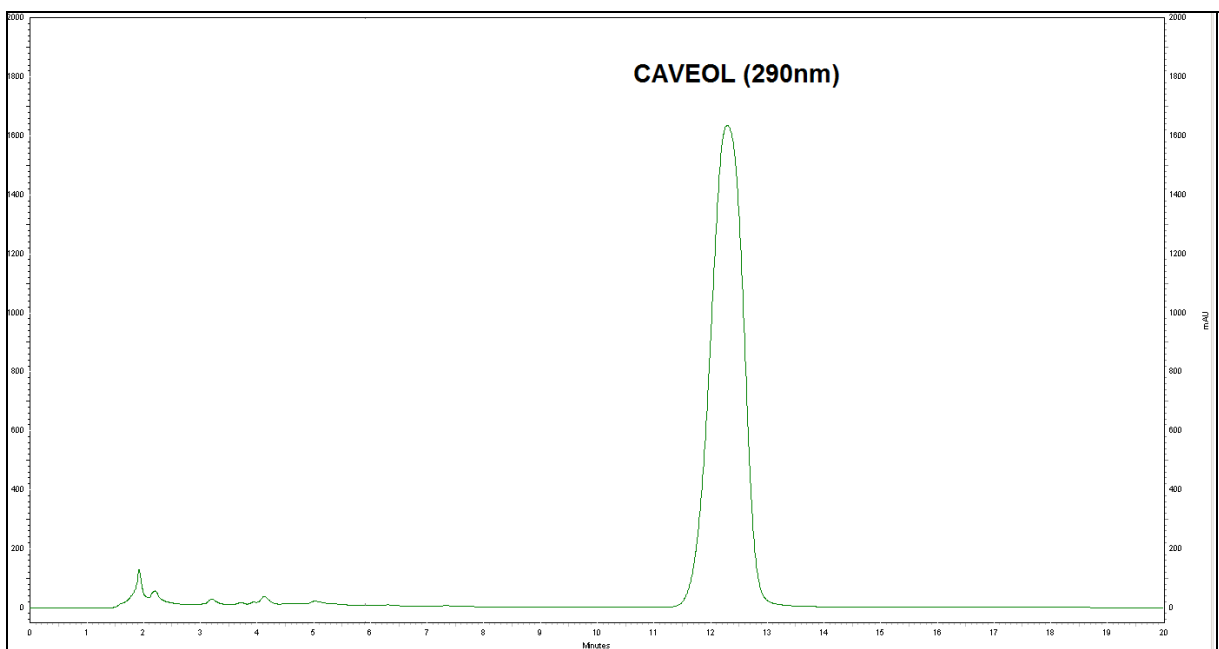
Figura 1: Via de Biossíntese dos CGAs e as enzimas. Modificado de MELO (2005).

ESPÉCIE	E				I	E				I	E				I	I	I	E						
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
<i>C. arabica</i>	T	T	T	T	G	A	T	T	A	A	G	C	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	
<i>C. arabica</i>	T	T	T	T	G	A	T	T	A	A	G	C	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	
<i>C. arabica</i>	T	T	T	T	G	A	T	T	A	A	G	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	
<i>C. arabica</i>	NA	NA	NA	NA	G	A	T	T	A	A	G	C	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	
<i>C. arabica</i>	NA	NA	NA	NA	G	A	T	T	A	A	G	C	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	
<i>C. arabica</i>	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	G	C	C	C	A	NA	NA	NA	NA	NA	NA	NA	NA	
<i>C. arabica</i>	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	C	C	A	C	C	T	G	C	A	T	T	G	
<i>C. arabica</i>	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	C	C	A	C	C	T	G	C	A	T	T	G	
<i>C. canephora</i>	C	C	G	C	C	C	A	C	G	A	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	
<i>C. arabica</i>	C	C	G	C	C	C	A	C	G	A	A	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	
<i>C. arabica</i>	C	C	G	C	C	C	A	C	G	G	A	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	
<i>C. arabica</i>	C	C	G	C	C	A	A	C	G	A	A	T	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	
<i>C. canephora</i>	NA	NA	NA	NA	C	C	A	C	G	A	A	T	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	
<i>C. arabica</i>	NA	NA	NA	NA	NA	NA	A	C	G	G	A	T	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	
<i>C. arabica</i>	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	G	G	G	C	C	C	A	T	C	NA	NA	
<i>C. arabica</i>	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	G	G	G	C	C	C	A	T	C	C	T	
<i>C. canephora</i>	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	G	G	T	T	A	T	C	C	T	
<i>C. canephora</i>	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	G	C	C	C	A	T	C	C	T	
<i>C. canephora</i>	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	G	G	T	T	A	T	C	C	T	
<i>C. arabica</i>	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	C	T	T

Figura 2: Exemplo de Identificação e Caracterização de Polimorfismos Intra-espécies e Inter-espécies (Contig4 da 4CL). E: polimorfismos entre grupos de *C. arabica*. I: polimorfismos dentro dos grupos de *C. arabica*, indicados com seta azul. NA: ausência de alelos. A: adenina. C: citosina. G: guanina. T: timina. Grupo 1: grupo de *C. arabica* semelhante à *C. eugenioides*. Grupo 2: grupo de *C. arabica* semelhante à *C. canephora* e para se observar melhor essa homologia, inseriu-se no grupo 2 algumas seqüências de *C. canephora*.

**ANEXO E****Cromatograma da quantificação de Cafestol**

Cromatograma da quantificação de Cafestol (pico em 220nm) em perisperma de frutos de 120DAF (janeiro) por CLAE. Eixo X: Absorbância. Eixo Y: Tempo de retenção em minutos.



Cromatograma da quantificação de Caveol (pico em 290nm) em perisperma de frutos de 120DAF (janeiro) por CLAE. Eixo X: Absorbância (mAU). Eixo Y: Tempo de retenção em minutos.

## ANEXO F

Gráfico de porcentagem de similaridade entre as sequências nucleotídicas dos 7 genes candidatos.

		1	2	3	4	5	6	7
Contig19	1		45,47	33,18	24,81	25,76	19,17	23,68
Contig28	2	45,47		34,60	24,33	27,18	19,03	23,84
Contig37	3	33,18	34,60		26,86	27,58	17,16	20,37
Contig25	4	24,81	24,33	26,86		41,90	16,82	20,55
Contig33	5	25,76	27,18	27,58	41,90		18,23	21,82
Contig22	6	19,17	19,03	17,16	16,82	18,23		39,53
Contig23	7	23,68	23,84	20,37	20,55	21,82	39,53	

1= Contig19. 2= Contig28. 3=Contig37. 4=Contig25. 5=Contig33. 6=Contig22. 7=Contig23.  
Software CLC Workbench.