UNIVERSIDADE ESTADUAL DE CAMPINAS

INSTITUTO DE BIOLOGIA

RAPHAEL RICON DE OLIVEIRA

"ESTUDO DA EXPRESSÃO DE GENES MADS-BOX DURANTE O DESENVOLVIMENTO FLORAL EM Coffea arabica L."

Este exemplar corresponde à redação final
da tese defendida pelo(a) candidato (a)
happen have be a linera
Asial
a aprovada pela Comissão Julgadora.
An on the second metal and an

Dissertação apresentada ao Instituto de Biologia para obtenção do Título de Mestre em Biologia Vegetal.

Orientador: Prof. Dr. Marcelo Carnier Dornelas

Campinas, 2011

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA DO INSTITUTO DE BIOLOGIA – UNICAMP

OL43e	Oliveira, Raphael Ricon de Estudo da expressão de genes MADS-box durante o desenvolvimento floral em <i>Coffea arabica</i> L. / Raphael Ricon de Oliveira. – Campinas, SP: [s.n.], 2011.
	Orientador: Marcelo Carnier Dornelas. Dissertação (mestrado) – Universidade Estadual de Campinas, Instituto de Biologia.
	1. Gene MADS-box. 2. Desenvolvimento floral. 3. <i>Coffea arabica</i> . 4. Reação em cadeia da polimerase via transcriptase reversa. 5. Hibridização <i>in situ</i> . I. Dornelas, Marcelo Carnier. II. Universidade Estadual de Campinas. Instituto de Biologia. III. Título.

Título em inglês: Study of the MADS-box genes expression during floral development in *Coffea arabica* L.

Palavras-chave em inglês: MADS-box gene; Floral development; *Coffea arabica*; Reverse transcriptase polymerase chain reaction; *In situ* hybridization.

Área de concentração: Biologia Vegetal.

Titulação: Mestre em Biologia Vegetal.

Banca examinadora: Marcelo Carnier Dornelas, Paulo Mazzafera, Carlos Augusto Colombo. **Data da defesa:** 21/02/2011.

Programa de Pós-Graduação: Biologia Vegetal.

Campinas, 21 de fevereiro de 2011.

BANCA EXAMINADORA

Prof. Dr. Marcelo Carnier Dornelas (Orientador)

Prof. Dr. Paulo Mazzafera

Prof. Dr. Carlos Augusto Colombo

Prof(a). Dr(a). Sandra Maria Carmello-Guerreiro

Prof(a). Dr(a). Adriana Pinheiro Martinelli

Assinatura

Assinatura

Assinatura Assinatura (M. C. V

Assinatura

Aos meus avós Aurélia (*in memorian*) e Manuel Aos meus pais Maria do Carmo e Amauri Aos meus irmãos Leandro, Adriano e Fabrício Dedico

AGRADECIMENTOS

A Deus e aos meus guias espirituais, forças ainda incompreendidas que nos auxiliam na jornada evolutiva, confortando e ajudando a erguer a cabeça nos momentos mais difíceis.

Aos meus pais pelo amor incondicional, educação e apoio em todas as minhas decisões. Vocês são meus exemplos de garra, dedicação e perseverança!

Ao professor Marcelo Carnier Dornelas pela oportunidade, ensinamentos e orientação.

A FAPESP e ao CNPq por terem financiado este projeto.

Aos professores e funcionários do programa de Pós-Graduação e do Departamento de Biologia Vegetal pelo suporte, ajuda e ensinamentos.

Aos professores/pesquisadores: Elliot W. Kitajima (NAP/ESALQ-USP) pelas análises de microscopia de varredura, Siu Mui Tsai (Lab. de Genômica/CENA) pelo sequenciamento dos clones, Maria Bernadete Silvarolla (IAC) pelas coletas de material.

A banca examinadora pelo aceite do convite, sugestões e correções da dissertação.

Aos amigos e colegas de trabalho pela ajuda, companheirismo e por tantos bons momentos durante esses anos de convivência.

Ao amigo e professor Antonio Chalfun-Junior, pela amizade e por ter me iniciado na carreira científica. Sem seus conselhos essa história provavelmente teria sido bem diferente.

A minha amada noiva Thaís pelo amor, compreensão e companheirismo em absolutamente todos os momentos. Parceira de todas as aventuras e que me faz muito feliz.

Aos eternos amigos da Universidade Federal de Lavras pela amizade e incentivo.

Aos amigos de república por terem acolhido a mim (e meu cachorro Churrasco) na chegada a Campinas, pela amizade e por tantos momentos de diversão.

A todos que de alguma forma contribuíram para realização deste trabalho.

v

RESUMO	vii
SUMMARY	viii
1. INTRODUÇÃO	1
2. OBJETIVOS	8
3. MATERIAIS E MÉTODOS	9
Identificação de sequências MADS-box	9
Classificação das sequências MADS-box	10
Identificação de domínios conservados	10
Material vegetal	11
Microscopia óptica e de varredura	12
RT-PCR	13
Hibridização in situ	14
4. RESULTADOS	. 17
Identificação e classificação de sequências	. 17
Caracterização morfo-anatômica do desenvolvimento floral	22
Perfis de expressão dos genes MADS-box	28
5. DISCUSSÃO	39
Características anatômicas	39
Os genes MADS-box em Coffea L.	. 41
O modelo ABC em Coffea arabica L	51
6. CONCLUSÕES	57
7. REFERÊNCIAS BIBLIOGRÁFICAS	59
8. APÊNDICE	66

ÍNDICE

RESUMO

A família dos genes MADS-box codifica fatores de transcrição que atuam como reguladores importantes em muitas etapas no desenvolvimento de diversos organismos. Em plantas, estes genes estão envolvidos na determinação da identidade dos meristemas reprodutivos e dos órgãos florais, bem como no controle de diversos processos durante o desenvolvimento. O presente trabalho teve como objetivo estudar o padrão de expressão dos prováveis ortólogos dos genes do modelo ABC (APETALA1, APETALA3, PISTILATA e AGAMOUS de Arabidopsis thaliana, e do gene TM6 de Solanum lycopersicum) em Coffea arabica L. Estes genes pertencem à família MADS-box e estão relacionados à determinação da identidade dos órgãos florais na planta-modelo A. thaliana. A partir do banco de dados de sequências expressas de cafeeiro (CAFEST), foram identificados 23 possíveis homólogos de genes MADS-box em cafeeiro. Perfis de expressão por RT-PCR indicaram que a maioria destes genes são expressos em flor e fruto. A análise dos dados gerados pelo uso de microscopia óptica e de varredura permitiu estabelecer uma sequência de desenvolvimento para estabelecimento dos órgãos florais em cafeeiro, facilitando a identificação dos locais de expressão dos ortólogos do modelo ABC pela técnica de hibridização in situ. Sendo C. arabica uma espécie relativamente recente e com características peculiares, foi proposto um mecanismo de atuação dos genes do modelo ABC. Dessa forma, os resultados obtidos contribuem para a compreensão do estabelecimento dos órgãos florais em C. arabica. Adicionalmente, pela caracterização de um número elevado de genes da família MADS, foram genes identificados outros potencialmente envolvidos em outros processos de desenvolvimento, que futuramente poderão ser utilizados para incremento da indústria cafeeira.

vii

SUMMARY

The MADS-box gene family encodes transcription factors that act as key regulators in many steps in the development of various organisms. In plants, these genes are involved in determining the identity of reproductive meristems and floral organs as well as in controlling several processes during development. This work aimed to study the expression patterns of putative orthologs of the ABC model genes (APETALA1, APETALA3, AGAMOUS and PISTILATA from Arabidopsis thaliana, and TM6 from Solanum Lycopersicum) in Coffea arabica L. These genes belong to the MADS-box family and are related to the determination of floral organ identity in the model plant A. thaliana. From the CAFEST database of expressed sequence tags, 23 MADS-box gene sequences were identified in coffee. Expression profiles of these genes, determined by RT-PCR, indicated that most of these genes are expressed in flowers and fruits. The analysis of data from optical microscopy and scanning electron microscopy allowed the establishment of a developmental sequence for the establishment of floral organ, facilitating the characterization of the spatial expression patterns of orthologs of the ABC genes by *in situ* hybridization. A diversified role of conserved genes of the ABC model was proposed for the relatively recent and peculiar specie that is C. arabica. The obtained results aid the understanding of the establishment of floral organs in C. Arabica. Additionally, as many other coffee MADS-box genes were also characterized, other genes, potentially involved in other developmental processes that could be of interest to the industry in the future were also identified.

1. INTRODUÇÃO

Os genes MADS-box

Genes MADS-box constituem uma família de fatores de transcrição envolvidos no controle dos principais aspectos do ciclo de vida das plantas terrestres (Gramzow & Theissen, 2010). Membros dessa família têm uma região (domínio MADS) de 58 aminoácidos altamente conservados, que ativam processos de transcrição ligando-se a elementos de reconhecimento - chamados CarG boxes (CC (A/T)₆ GG) encontrados em promotores de genes alvo no DNA (Riechmann *et al.*, 1996). O termo MADS-box refere-se às iniciais dos primeiros quatro membros descritos, *MINICHROMOSOME MAINTENANCE 1 (MCM1)* de *Saccharomyces cerevisiae*, *AGAMOUS (AG)* de *Arabidopsis thaliana*, *DEFICIENS (DEF)* de *Antirrhinum majus* e *SERUM RESPONSE FACTOR (SRF)* de *Homo sapiens* (Schwarz-Sommer *et al.*, 1990). Estudos recentes indicam que genes MADS-box originaram de uma sequência de DNA (Gramzow *et al.*, 2010). Uma das cópias acumulou mudanças nas sequências originando um domínio com maior especificidade na ligação ao DNA, o domínio MADS (Gramzow *et al.*, 2010).

Como um resultado da subsequente duplicação e divergência do gene, dois principais tipos de genes MADS-box são reconhecidos - Tipo I (SRF-like) e tipo II (MEF2-like) estabelecidos no ancestral mais recente de eucariotos (Alvarez-Buylla *et al.*, 2000a). Esses dois tipos são distinguidos pela sequência de DNA, especificidade do local de ligação ao DNA e a quantidade de transcritos que eles induzem (Messenguy & Dubois, 2003). Os dois tipos são distinguidos pelos domínios característicos da estrutura de suas proteínas, sendo que fatores de transcrição MADS do tipo II de plantas apresentam domínio keratin-like (K) e são geralmente conhecidos como proteínas MIKC, devido à presença de quatro domínios característicos: MADS (M) altamente conservado, Intervening (I) correspondente a uma região interna que se conecta ao domínio Keratin-like (K), por sua vez, responsável por interações proteína-proteína, seguindo-se então uma porção Carboxy-terminal (C), envolvida na ativação da transcrição (Parenicova *et al.*, 2003; Becker & Theissen, 2003). As proteínas do tipo I não possuem o domínio K e são subdivididas em três subfamílias, M α , M β e M γ , podendo ou não ter domínios característicos (De Bodt *et al.*, 2003).

A família de genes MADS-box pode ser dividida em diversos clados ancestrais revelados por uma reconstrução filogenética entre diversas espécies de Angiospermas (Becker & Theissen, 2003), que demonstra relações gênicas ortólogas extremamente conservadas ou parálogas originadas de eventos recentes de duplicação retomando os seguintes grupos: AG, AGL2, AGL6, AGL12, AGL15, AGL17, DEF, FLC, GGM13, GLO, SQUA, STMADS11 e TM3. O grande número de genes MADS-box encontrados em alguns grupos de plantas, em contraste com o número reduzido em protistas, animais e fungos, indica a importância desses fatores de transcrição contribuindo para diversos processos biológicos durante a evolução das plantas (Gramzow & Theissen, 2010). Em *A. thaliana*, por exemplo, são encontrados 107 genes, em *Oriza sativa* (arroz) 71 e em *Populus trichocarpa* 105 (Parenicova *et al.*, 2003; Leseberg *et al.*, 2006), que estão envolvidos em diversos processos de desenvolvimento da planta, nem sempre relacionados ao desenvolvimento reprodutivo (Zhang & Forde, 1998; Alvarez-Buylla *et al.*, 2000b).

O modelo ABCE

A grande maioria dos genes relacionados à determinação da identidade dos meristemas e dos órgãos florais pertence à família MADS-box (Krizek & Fletcher, 2005). A partir de estudos envolvendo principalmente mutantes MADS-box em *A. thaliana* e *A. majus*, foi proposto um modelo composto por três fatores: A, B e C, para explicar a determinação da identidade dos órgãos florais (Coen & Meyerowitz, 1991). Esses fatores, representados pelo produto de genes homeóticos específicos, atuam de forma combinatória determinando a formação dos quatro verticilos florais (Theissen & Saedler, 2001; Honma & Goto, 2001). Segundo Coen & Meyerowitz (1991), a função A conferida pelos genes *APETALA1 (AP1)* e *APETALA2* em *A. thaliana* e seus respectivos ortólogos *LIPLESS1* e *LIPLESS2* em *A. majus*, especificam a identidade das sépalas no primeiro verticilo. Essa mesma função combinada com a função B, conferida pelos genes *APETALA3 (AP3)* e *PISTILATA (PI)* em *A. thaliana* e *DEFICIENS* e *GLOBOSA* em *A. majus*, especifica a identidade de pétalas no segundo verticilo. A função B combinada com a função C, conferida pelos genes *AGAMOUS (AG)* em *A. thaliana* e *PLENA (PLE)* e *FARINELLI (FAR)* em *A. majus*, especifica identidade de estames no terceiro verticilo. E por fim, a função C expressa sozinha no quarto verticilo especifica identidade do carpelo, além de conferir determinação do meristema floral.

O programa básico de desenvolvimento floral abrangido pelo modelo ABC parece ser amplamente conservado entre as espécies de plantas estudadas (Krizek & Fletcher, 2005). Porém, diversos estudos comparando espécies filogeneticamente distantes mostram que eventos de duplicação seguidos de funcionalização são constantes e podem tomar caminhos diferentes na evolução, sendo esse o principal mecanismo na origem da diversidade morfológica (Litt & Kramer, 2010). Exemplo disso ocorre com uma duplicação de genes de função B na base das eudicotiledôneas, onde é sugerido que tenha dado origem a um perianto diferenciado de padrão claramente observado entre essas espécies, ao contrário de angiospermas mais basais, ausentes dessa cópia gênica e apresentando diferentes morfologias florais com perianto indiferenciado (Theissen & Melzer, 2007). A descoberta de que a superexpressão de genes do modelo ABC em folhas não é suficiente para determinação da identidade do órgão floral indicou a existência de outro fator também essencial a essa identidade (Pelaz *et al.*, 2000). Estudos envolvendo a classe de genes *SEPALLATA* identificaram quatro parálogos em *A. thaliana* com funções redundantes e essenciais à formação de complexos específicos em cada verticilo, sendo chamados de genes de função E (Pelaz *et al.*, 2000; Honma & Goto, 2001). Estudos recentes indicam que esses genes funcionam como verdadeiros conectores na formação de complexos de ordem superior, interagindo com diferentes fatores de transcrição e desencadeando uma série de processos coordenados durante o desenvolvimento (Immink *et al.*, 2009).

A existência de genes MADS-box em gimnospermas, samambaias e musgos, os quais não formam flores ou frutos, demonstra que a função desses genes em plantas não é restrita ao desenvolvimento de órgãos reprodutivos (Münster *et al.*, 2002b). De forma concordante, estudos têm demonstrado que os genes MADS-box estão envolvidos em muitos aspectos no desenvolvimento de plantas (Zhang & Forde, 1998; Alvarez-Buylla *et al.*, 2000b).

Características de Coffea arabica L.

O gênero *Coffea* L. pertence à família Rubiaceae e conta com 103 espécies identificadas (Davis *et al.*, 2006). A grande maioria das espécies conhecidas são diplóides (2n=2x=22) e munidas de sistemas de auto-incompatibilidade reprodutiva. Uma das exceções é a espécie *C. arabica*, alotetraplóide (2n=4x=44) com taxa de auto-fecundação em torno de 90% (Grassias & Kammacher, 1975; Charrier & Berthaud, 1985). Híbridos poliplóides férteis podem ter surgido de diferentes formas no gênero *Coffea* (Harlan & deWet, 1975), e embora não esteja claro o modo como ocorreu na espécie *C. arabica*, estudos indicam que trata-se de um anfidiplóide formado da hibridização dos genomas de duas espécies próximas: *Coffea*

eugenioides, parente feminino, e Coffea canephora Pierre (Lashermes et al., 1999; Maurin et al., 2007).

As espécies do gênero *Coffea* são árvores e arbustos, sendo o cafeeiro caracterizado pelo dimorfismo dos ramos, classificados como ortotrópicos (crescimento vertical e simetria radial) e plagiotrópicos ou laterais (crescimento horizontal ou obliquo, com simetria bilateral) (Carvalho et al., 1950). Em ramos ortotrópicos são relatados dois tipos de gemas axilares que diferem basicamente pela estrutura que originam: gemas cabeça de série, que dão origem aos ramos plagiotrópicos, e gemas seriadas, que podem formar ramos ortotrópicos de substituição (Moens, 1963). Nos ramos plagiotrópicos existem apenas as gemas seriadas (Wormer & Gituanja, 1970b) que podem formar ramos plagiotrópicos de ordem superior e inflorescências ou permanecerem indiferenciadas (Carvalho et al., 1950). Dedecca (1957) numa descrição morfológica em C. arabica, relata que as inflorescências desenvolvem-se num eixo curto terminando em flor. Nestes novos eixos surgem pares de brácteas cruzadas, que podem desenvolver outras séries descendentes de gemas. Os eixos que suportam as flores são sempre muito curtos e a inflorescência em conjunto assume aspecto compacto de glomérulo. Os primórdios florais aparecem na axila formada pelas brácteas com o eixo da inflorescência. A flor do cafeeiro apresenta um cálice de cinco pequenos segmentos. A corola é de cor branca, constituída de cinco pétalas unidas entre si até a metade formando um tubo. Os cinco estames são epipétalos e inserem-se sobre o tubo da corola. Cada estame consiste em um filamento cilíndrico e curto inserido no setor mediano de uma antera bilocular com quatro sacos polínicos. O gineceu é formado por dois carpelos unidos pelo estilete que se bifurcam em dois lobos estigmáticos, sendo o ovário ínfero bilocular contendo um óvulo por lóculo.

Fisiologia do florescimento em Coffea arabica

O ciclo fenológico do cafeeiro arabica apresenta uma sucessão de fases vegetativas e reprodutivas que ocorrem a cada dois anos, diferentemente da maioria das plantas cujas inflorescências surgem na primavera e os frutos formam-se no mesmo ano fenológico (Camargo, 1985). O primeiro ano fenológico é caracterizado pela formação dos ramos vegetativos com gemas axilares nos nós, posteriormente induzidas a um estado reprodutivo (Gouveia, 1984). Em seguida, as gemas florais amadurecem, entram em dormência e tornam-se aptas à antese, que ocorre cerca de dez dias após uma ação indutória causada principalmente por chuva ou irrigação (Camargo, 1985). O segundo ano fenológico inicia-se com a florada, seguida da formação dos chumbinhos (frutos imaturos com cerca de 3-4 mm de diâmetro) e da expansão dos grãos. Finalmente, ocorre a granação dos frutos e a fase de maturação (Camargo & Camargo, 2001). Nas condições tropicais do Brasil, cada ano fenológico geralmente inicia-se em setembro e termina em agosto do ano seguinte.

São relatados três períodos de quiescência antes da antese nas gemas reprodutivamente induzidas (Wormer & Gituanja, 1970a). O primeiro ocorre imediatamente após a diferenciação e o segundo quando as gemas já se encontram entumescidas, volumosas e cobertas por secreção. O terceiro ocorre quando os botões florais individuais já se encontram totalmente formados. Apesar do estresse hídrico não ser uma condição necessária para ocorrência da florada, o período de seca precedente parece concentrar as floradas (por impedir antese prematura das gemas florais mais adiantadas) e promover a aceleração da etapa final de diferenciação das gemas florais (Gouveia, 1984). O período de florescimento é caracterizado por sucessivas floradas, havendo sempre uma florada principal, em consequência do processo assincrônico de diferenciação e crescimento das gemas reprodutivas e da distribuição das chuvas no período (Frederico & Maestri, 1970).

Morais *et al.* (2008) propuseram uma escala fenológica detalhada da fase reprodutiva de *C. arabica* dividindo seu desenvolvimento em quatro grandes fases: desenvolvimento da gema floral (G), floração (FL), frutificação (F) e maturação (M). As fases G e F foram subdivididas tomando-se como critério o tamanho das gemas e frutos, respectivamente. Para a fase M o critério adotado foi a coloração dos frutos. No entanto, a duração das fases e o tempo necessário para a completa maturação sofrem influência de condições climáticas (Kumar, 1979) e da constituição genética do cafeeiro (Söndahl & Sharp, 1979).

Com base nessas informações, torna-se provável que o florescimento e o desenvolvimento de frutos do cafeeiro estejam associados às variações edafoclimáticas, principalmente àquelas associadas à alteração no potencial hídrico das plantas (chuva, irrigação após período seco e rápido decréscimo de temperatura). Porém, do ponto de vista molecular, o mecanismo envolvido na percepção desses estímulos pelas plantas é desconhecido. Igualmente, pouco se sabe a respeito da conservação dos fatores moleculares responsáveis pela determinação da identidade dos órgãos florais em cafeeiro.

Justificativa

Dentre as diversas espécies pertencentes ao gênero *Coffea*, *C. arabica* é a que produz bebida de melhor qualidade, respondendo por uma produção em torno de 64% do total mundial (Fasio & Silva, 2007). O processo de florescimento depende da expressão equilibrada de uma rede complexa de genes, que é regulada por fatores endógenos e ambientais. Um dos problemas na qualidade do café consiste no florescimento sequencial dos botões florais e, consequentemente, desuniformidade na maturação dos frutos, o que dificulta a colheita e prejudica a qualidade dos grãos. No momento, a complexidade da rede de interações moleculares responsável pelo controle do florescimento está em grande parte baseada em estudos realizados em *A. thaliana* (Izawa *et al.*, 2003). Contudo, extensivos esforços continuam sendo feitos para descrever vias relacionadas em outras espécies que auxiliem no entendimento do mecanismo básico de desenvolvimento floral (Benlloch *et al.*, 2007). A aplicação desses conceitos à espécie *C. arabica* seria de grande interesse para elucidação de seu complexo processo de desenvolvimento reprodutivo, permitindo o incremento de práticas agrícolas (por exemplo, obtenção de plantas macho- e/ou fêmea-estéreis e aumento do número de verticilos produzindo um determinado órgão floral – ver revisão Krizek & Fletcher, 2005), além de fornecer dados moleculares inéditos de como ocorre a iniciação e o estabelecimento dos órgãos florais nessa espécie.

2. OBJETIVOS

O projeto teve como objetivo identificar em *C. arabica* os genes ortólogos aos genes do modelo ABC - *APETALA1 (AP1)*, *APETALA3 (AP3)*, *PISTILATA (PI)*, *AGAMOUS (AG)* e *TOMATO MADS 6 (TM6)* - e estudar seus padrões de expressão relacionando-os às alterações anatômicas ocorridas durante o desenvolvimento floral. Dessa forma, inferir suas prováveis funções e a organização do modelo ABC no cafeeiro.

Os objetivos específicos foram: obter sequências MADS-box completas a partir do banco de ESTs CAFEST; classificá-las e indicar seus prováveis ortólogos nos subgrupos relatados; estabelecer as fases do desenvolvimento reprodutivo por microscopia óptica e eletrônica de varredura; determinar por RT-PCR os padrões de expressão desses genes nos diferentes tecidos e fases de desenvolvimento reprodutivo; determinar o padrão espacial de

expressão dos ortólogos do modelo ABC durante o desenvolvimento floral pela técnica de hibridização *in situ*.

3. MATERIAIS E MÉTODOS

Identificação de sequências MADS-box

As sequências ESTs de *C. arabica* homólogas aos genes da família MADS-box foram obtidas no banco de dados CAFEST por meio de duas ferramentas disponíveis no site do Projeto Genoma Brasileiro do Café (Vieira *et al.*, 2006): busca por palavra-chave e busca por similaridade - BLAST (Altschul *et al.*, 1997). Para esta última, foi utilizada uma sequência consenso da região MADS gerada pelo programa COBBLER (Henikoff & Henikoff, 1997) a partir de sequências publicadas de genes MADS-box. Foram selecionados todos os reads que apresentaram similaridade significante (e-value > 10^{-5}) e agrupados pelo programa CAP3 (Huang & Madan, 1999), formando os chamados EST-contigs. Para verificação da presença do domínio conservado MADS as sequências foram inspecionadas pelo programa InterProScan (The InterPro Consortium, 1999) e sendo o BLAST um método heurístico de procura, todos os EST-contigs validados foram usados para uma terceira busca por similaridade visando-se encontrar novos reads e remontar clusters incompletos. Esse processo foi repetido até que não fosse encontrado mais nenhum novo read significante.

Os clones bacterianos mais representativos de cada EST-contig foram solicitados ao Banco de Clones (BCCC, Jaboticabal, SP). Foi feito o isolamento de colônias em câmara de fluxo esterilizada e crescimento dessas bactérias em meio LB ágar seletivo (Ampicilina 100 μ g/ml), por 12 horas a 37°C, seguido de seu cultivo em meio LB líquido também seletivo, em agitador (200 rpm) pelo mesmo período e temperatura. Com essa cultura de células bacterianas foram realizados dois procedimentos: estoque e armazenamento dos clones

bacterianos (1:1 em glicerina pura) a -80°C e, extração dos plasmídeos utilizando-se o kit PureLinkTM Quick Plasmid Miniprep (Invitrogen). Os plasmídeos extraídos foram então precipitados em NaAc 3M pH 5,2 e EtOH absoluto gelado, seguido de centrifugação e secagem, e enviados à plataforma de sequenciamento do Laboratório de Genômica do Centro de Energia Nuclear na Agricultura (CENA/USP, Piracicaba, SP).

Para algumas sequências contendo longas caudas poli-A, foi utilizada a técnica de primer walking, na qual primers internos são utilizados no sequenciamento. Esses primers foram desenhados em regiões confiáveis das sequências e tinham em geral 20 bases e temperatura de anelamento de 60°C.

Classificação das sequências MADS-box

A classificação das sequências MADS encontradas no CAFEST foi feita por uma análise filogenética comparando-os aos genes de *A. thaliana* (Parenicova *et al.*, 2003). Todas as sequência, tanto as de *C. arabica* quanto as de *A. thaliana*, foram alinhadas utilizando-se o programa ClustalW com os parâmetros padrões (Thompson *et al.*, 1994). A árvore final foi obtida utilizando-se o programa MEGA4 (Tamura *et al.*, 2007) com o algoritmo Neighborjoining (Saitou & Nei, 1987) e validadas pelo teste probabilístico de bootstraps (Sitnikova *et al.*, 1995).

Identificação de domínios conservados

Para descobrir domínios conservados presentes nas sequências MADS selecionadas no CAFEST, utilizou-se o programa MEME (*Multiple Expectation Minimization for Motif Elicitation*) versão 4.4.0 (Bailey & Elkan, 1994). Os parâmetros utilizados foram: número de repetições qualquer, máximo número de motifs 4 e amplitude ótima entre 6 e 200. Os motivos foram anotados quanto aos domínios funcionais de proteínas por meio do software SMART (*Simple Motif Architecture Research Tool*) versão 5.0 (Schultz *et al.*, 1998; Letunic *et al.*, 2002), capaz de reconhecer mais de 500 domínios presentes em diferentes tipos de proteína.

Material vegetal

Foram utilizadas como fonte de tecidos plantas de *C. arabica* cv. Mundo Novo em idade reprodutiva já estabelecidas no campo experimental do Departamento de Biologia Vegetal da Unicamp e do Instituto Agronômico de Campinas (IAC), além de sementes da mesma cultivar germinadas e cultivadas em hidroponia. Os tecidos vegetais escolhidos para as diversas análises foram: raiz, entrenó de ramos plagiotrópicos (caule jovem), folhas jovens, meristema apical de ramos plagiotrópicos (ápice ou ponteiro) e diferentes estádios de desenvolvimento de meristemas axilares de ramos plagiotrópicos (gemas axilares), botões florais e frutos.

Para coleta de raiz, sementes foram germinadas em papel filtro e água destilada mantidos no escuro a uma temperatura de 27°C até o desenvolvimento e abertura dos cotilédones (período de 45 dias). As plântulas foram translocadas para hidroponia oxigenada em solução nutritiva de Hoagland modificada trocada quinzenalmente, e mantidas por três meses. Os meristemas axilares foram coletados mensalmente na altura do nó de ramos plagiotrópicos durante o período de dezembro de 2008 a junho de 2009. Os botões florais e frutos foram selecionados quanto ao critério de tamanho, de acordo com Morais *et al.* (2008). Para cada tecido foram coletados respectivamente seis diferentes fases: G2, G3, G4, G5, G6 e FL, no período de maio a agosto de 2009 e F1, F2, F3, F4, F5 e F6, entre agosto e setembro de 2009.

Os tecidos destinados à microscopia óptica (MO), microscopia eletrônica de varredura (MEV) e hibridização *in situ* foram fixados em paraformaldeído 4% (p/v) por um período de

12-16 horas a 4°C e, posteriormente, submetidos a uma série de desidratação em álcool etílico (EtOH). Após atingirem a concentração de etanol absoluto, os tecidos foram armazenados a 4°C. Para os tecidos destinados a RT-PCR, ou seja, para extração de RNAs, foi feita a coleta diretamente em nitrogênio líquido (N₂) e armazenados a -80°C.

Microscopia óptica e de varredura

Para a microscopia óptica foram utilizados os materiais já fixados e armazenados em álcool absoluto, como descritos acima, e incluídos em resina plástica ativada (Historesin, Leica) seguindo-se uma série de inclusão nas seguintes proporções Resina/EtOH: 1:3, 1:1 e 3:1, com duração de 24 horas cada uma. Na última etapa o material ficou imerso por um período 48 horas em resina pura. Após a série de inclusão o material foi polimerizado em moldes plásticos seguindo as instruções do fabricante. Os blocos de resina polimerizada foram colados em pequenos cubos de madeira, sendo submetidos aos cortes histológicos seriados em micrótomo rotativo. Foram feitas secções de 5 µm e os cortes sobrepostos em água sobre lâmina, seguido de secagem em mesa térmica a 50°C. Os cortes foram então corados com azul de toluidina (0,05% p/v em fosfato de sódio bibásico anidro e ácido cítrico, pH 5) por 4 minutos e montados permanentemente com Entellan entre lâmina e lamínulas. As lâminas permanentes foram observadas e fotografadas no microscópio ZEISS modelo AXIOVERT 35 do Departamento de Biologia Vegetal da Unicamp.

Para a microscopia de varredura os mesmos materiais fixados e armazenados em álcool absoluto foram utilizados, porém foram secos ao ponto crítico. Após dissecação adicional das amostras e metalização com ouro coloidal (camada de 30nm), a observação foi feita em microscópio de varredura LEO 435 VP e as imagens digitalizadas foram obtidas utilizando-se o programa LEOUIF.

A extração de RNA total dos materiais vegetais descritos acima foi feita pelo método descrito por Rezaian & Krake (1987) com modificações para retirada de DNA genômico contaminante. Uma amostra de cada tecido foi bem macerada com pilão e nitrogênio líquido em almofariz e, a 2,5 g desse pó foi adicionado o tampão de extração (Tris-HCl 1M pH 8,3, NaClO₄ 9M, PEG 8000, H₂O DEPC, SDS 20% w/v, PVPP 8,5% v/v, 6 μl β-mercaptoetanol) sendo agitado rapidamente a temperatura ambiente por alguns minutos. A solução foi então centrifugada a 14000 rpm por 10 minutos a 4°C. Após coleta da fase sobrenadante, foi adicionado 800 µl de EtOH absoluto gelado, sendo mantidos a -20°C por 20 minutos para precipitação de ácidos nucléicos. Foi feita nova centrifugação a 8000 rpm por 15 min a 4°C e o pellet formado lavado com 300 µl de EtOH 70% gelado. Após repetição da última centrifugação, o pellet foi ressuspendido em 300 µl de Tris-EDTA (10 mM/5mM, pH8,0) e 1 μl de β-mercaptoetanol. O tubo foi agitado e adicionado 750 μl de fenol/clorofórmio. Seguiuse centrifugação a 10000 rpm por 10 min e coleta do sobrenadante. Foi adicionado 750 µl de clorofórmio/álcool isoamílico (24:1 v/v), agitado e novamente centrifugado a 10000 rpm por 5 min. A fase aquosa foi coletada e o RNA precipitado pela adição 50 µl de acetato de sódio $(3M, pH 5,2) = 750 \mu l$ de etanol absoluto gelado e incubados a -20°C por cerca de 1 hora. Foi feita mais uma centrifugação a 10000 rpm por 15 min a 4°C, e o pellet lavado com 750 µl de EtOH, que após nova cetrifugação a 8000 rpm por 5 min a 4°C e secagem ao ar livre, pode ser ressuspendido em H₂O DEPC. A extração dos RNA para os tecido pôde ser confirmada em gel de agarose.

Para síntese dos cDNAs, os RNAs extraídos de cada tecido foram quantificados em espectrofotômetro e após cálculo do volume necessário para uma quantidade de 2,5 µg de RNA, foram feitos os tratamentos com DNAse (DNA-free, Ambion) para retirada de

remanescentes de DNA genômico seguindo-se as instruções do fabricante. Em seguida, a primeira fita de cDNA foi sintetizada em reação com transcriptase reversa (Kit Superscript II, Invitrogen®) e primers oligo-dT. Estes constituíram os moldes para as reações de PCR.

Os primers de cada gene MADS selecionado no cafeeiro foram desenhados em regiões não conservadas das sequências utilizando-se o programa Primer3 (Rozen & Skaletsky, 2000), tendo os seguintes parâmetros solicitados: amplificação de produtos de PCR entre 380 e 450 pb, tamanho de 18 a 22 bases, melting temperature (Tm) em torno de 55°C e conteúdo GC entre 45-55% (para a sequência dos primers, ver APÊNDICE – Item 1). Foram selecionados primers com baixa energia livre de complementaridade (Δ G), após serem testados no programa Gene Runner 3.05 (Copyright © 1994 Hastings Software, Inc.) quanto à formação de dímeros e hairpins. Posteriormente, os primers foram testados em reações de PCR utilizando-se o próprio plasmídeo do respectivo clone como template (controle positivo).

O experimento para documentação dos perfis de expressão das sequências selecionadas, foi precedido pela normalização das quantidades de cDNA nas amostras de tecidos coletados. Para tal, foram desenhados primers específicos do gene constitutivo *GAPDH*, indicado como tendo expressão bastante estável em diversos tecidos de cafeeiro (Barsalobres-Cavallari *et al.*, 2009). Com os tecidos normalizados, as reações de PCR semi-quantitativas foram realizadas e documentadas digitalmente pelo programa Gel Doc (Syngene).

Hibridização in situ

Para síntese da sonda foram utilizados como moldes os clones contendo os respectivos cDNAs dos ortólogos dos genes do modelo ABC (AP1, AP3, TM6, PI e AG). Após crescimento das bactérias em meio LB ágar/ampicilina (100µg/ml), os plasmídeos foram

extraídos (kit miniprep Invitrogen®) e linearizados com enzima EcoRI numa reação que levou 75 μ l de miniprep em TE (Tris-EDTA), 10 μ l H₂O, 10 μ l tampão da enzima e 5 μ l de EcoRI, totalizando 100 μ l que foram mantidos a 37°C por 2 horas. Um segundo tipo de sonda foi feita com o mesmo procedimento linearizando-se os plasmídeos numa região mediana ao clone utilizando-se as enzimas SACI, HindIII e HindFI, especificas para cada clone. As sondas foram obtidas por transcrição *in vitro* dos plasmídeos linearizados com emprego da transcriptase SP6 em reação contendo uracila marcada com digoxigenina (DIG-UTP), segundo as instruções do fabricante do kit (Roche).

Os materiais já fixados e armazenados em álcool absoluto, conforme descrito anteriormente, foram transferidos para xilol seguindo-se uma série de proporções xilol/EtOH: 1:3, 1:1 e 3:1, com duração de 24 horas cada uma e, posteriormente mantidos por 3 dias em xilol puro. Foi feita a substituição do xilol por parafina adicionando-se ao material parafina histológica granulada e mantendo-o em estufa a 60°C. Para retirada total do xilol foram feitas quatro trocas de parafina líquida por dia durante quatro dias. O material totalmente incluso em parafina líquida foi então posicionado e solidificado em formas de papel e montados em cubos de madeira para os cortes histológicos em micrótomo. Os cortes seriados com 7µm de espessura foram colocados sobre H₂O (tratadas com DEPC) sobre lâmina previamente submetidas a um banho de organosilana (3-aminopropiltrietoxisilana, Pierce 2% em acetona), por 6 minutos. As lâminas foram secas em chapa elétrica a 50°C até e armazenadas a 4° até que fossem utilizadas para hibridização.

A hibridização *in situ* consistiu de três etapas: pré-hibridização/hibridização, póshibridização/imuno-detecção e visualização das lâminas. Para primeira etapa foi feita a remoção da parafina dos cortes com banhos de xilol:EtOH nas proporções 3:1, 1:1, 1:3 e álcool absoluto, sendo secos posteriormente ao ambiente. As lâminas foram então selecionadas em microscópio, submetidas ao tratamento de pré-hibridização com proteinase K (1µg/ml em 0,05M Tris HCl pH:7,5) durante 13 minutos e lavadas em H₂O DEPC. A hibridização foi realizada em tampão de hibridização (3ml Formamida deionizada, 132µl Tris-HCL 1M pH 8,0, 720µl NaCl 5M, 120µl EDTA, 144µl Denharts, 1,44 ml de Dextran Sulfato 50%) contendo 300-600ng de sonda. Foram colocados 300 µl dessa solução sobre as laminas, cobertas com parafilme e mantidas em câmara úmida a 42°C durante 16 horas.

Na segunda etapa, as lâminas foram imersas duas vezes em tampão SSC 2X (solução estoque 20X: 175,3g/L NaCl, 88,2g/L Citrato de Sódio, pH= 7) e duas vezes em tampão SSC 4X, mantidas a 42°C durante 30 minutos cada uma. Seguiram-se três lavagens em H₂O estéril e imuno-detecção com anticorpo anti-DIG conjugados à fosfatase alcalina. As lâminas foram deixadas por 5 minutos em tampão de detecção DB1 (100 mM Tris-HCL pH 8,0, 150 mM NaCl) e por 30 minutos em tampão de bloqueio, DB2 (DB1 acrescido de 1% Blocky agent, Roche). Foram colocados 300µl de solução de anticorpo (1:1000 em DB2) em cada lâmina. Posteriormente, as lâminas foram cobertas com parafilme e armazenadas em câmara úmida por 1 hora. As laminas foram imersas duas vezes por 15 minutos em DB1 e mais 5 minutos em DB3 (100mM Tris-HCL pH 8,0, 100mM NaCl e 50mM MgCl₂). A visualização do sinal foi obtida após aplicação de 300µl de substrato comercial de NBT/BCIP Plus Suppresor (Pierce) sobre cada lâmina, que foi em seguida coberta com lamínulas e mantidas por 12 horas no escuro. Foram feitas lâminas permanentes com Entellan e o material hibridizado foi observado e documentado em microscópio ZEISS modelo AXIOVERT 35.

4. RESULTADOS

Identificação e classificação de sequências

Os métodos de buscas por sequências no banco CAFEST permitiram a identificação de mais de 100 reads, que após processo de agrupamento e comparação com bancos de dados públicos, resultaram na formação de 23 EST-contigs relacionados à família de genes MADS-box. Os clones mais representativos dos 23 EST-contigs selecionados foram então resequenciados para confirmação (ver APÊNDICE, Item 2 - Lista de sequências). Esse procedimento revelou que dois contigs (20, 22) apresentavam a região MADS incompleta (cDNA clonado fragmentado a 5') e três (3, 16 e 20) não tinham seus clones disponíveis no Banco de Clones (BCCC, Jaboticabal, SP).

As análises filogenéticas incluindo as 23 sequências encontradas no CAFEST e as sequências publicadas de *A. thaliana* (Parenicova *et al.* 2003) permitiram a reconstrução dos grupos relatados para a família MADS-box (Theissen *et al.*, 2000), permitindo a classificação das mesmas nos diferentes subclados bem suportadas pelo teste de bootstraps (Figura 1). Todas as sequências foram classificadas como genes do tipo II de proteínas MADS-box (Alvarez-Buylla *et al.*, 2000), a grande maioria apresentou estrutura MIKC típica com os dominios característicos MADS e K-box bastante conservados (Figura 2), e apenas 7-CaMd apresentou uma estrutura mais similar ao grupo Mδ.



Figura 1 - Árvore filogenética relacionando sequências MADS de C. arabica (quadrado) e genes MIKC-type de A. thaliana publicados (entre parênteses, os respectivos números de acesso).

(MIKC-like)



Figura 2 – Representação gráfica dos "motifs" de agrupamento entre os 23 EST-contigs completos encontrados no CAFEST (Bailey & Elkan, 1994). Os motifs indicados acima foram anotados e estão representados pela figura do topo. A cor azul-clara representa a porção mais conservada do domínio MADS. A cor vermelha representa parte do domínio I. As cores azul-escuro e rosa representam dois "motifs" conservados do domínio K. A linha cinza ao final das sequências indica a porção C-terminal, não conservada.

Dessa forma, foram indicados os prováveis ortólogos em *C. arabica* para os diferentes genes MADS-box descritos em *A. thaliana* e de particular interesse para este trabalho: os candidatos ortólogos do modelo ABC (Função A: sequências 13 e 15, função B sequências 5, 10 e 14, e função C sequências 12 e 22 (Figura 1).

Tendo em vista a considerável distância filogenética entre as famílias de *A. thaliana e C. arabica*, e as diversas duplicações relatadas ao longo da evolução da família MADS-box, análises filogenéticas complementares, incluindo sequências de espécies mais próximas ao cafeeiro, foram realizadas para separar esses subgrupos e melhor caracterizar os verdadeiros ortólogos de genes do modelo ABC. Para a função A, as análises confirmaram a ortologia da sequência 13 à *AP1*, sendo nomeada *Coffea arabica APETALA1* (*CaAP1*, Figura 3A). Para a função B, a sequência 5 apresentou ortologia à *AP3* (*Coffea arabica APETALA3, CaAP3*), a sequência 10 foi classificada como ortóloga ao gene *PI* (*Coffea arabica PISTILLATA, CaP1*), e a sequência 14, ortóloga à *TM6* (*Coffea arabica TOMATO MADS 6, CaTM6*; Figura 3B). Por fim, para a função C o ortólogo ao gene *AG* foi a sequência 12 (*Coffea arabica AGAMOUS, CaAG*; Figura 3C).

Como cada EST-contig foi oriundo de montagens incluindo diversos reads, diferenças alélicas (SNPs ou indels, deleções e translocações de bases) e splicing alternativos, puderam ser observadas e encontram-se anotadas na lista de sequências (APÊNDICE – Item 2).







Figura 3 - Árvores específicas comparando sequências MADS de *C. arabica* (quadrado) e genes publicados das respectivas famílias de diversas espécies (números de acesso e espécie estão acompanhados). **A** – Subclado SQUA (função A), **B** – Subclado Def+GLO (função B), e **C** – Subclado AGAMOUS (função C).

Caracterização morfo-anatômica do desenvolvimento floral

As análises feitas em microscopia óptica e eletrônica de varredura permitiram estabelecer as etapas do desenvolvimento do meristema vegetativo durante sua transição para meristemas de inflorescência, seguindo-se a formação dos meristemas florais e desenvolvimento dos primórdios dos órgãos florais. Meristemas vegetativos oriundos de ramos ortotrópicos são menores, mais achatados e compactos quando comparados com gemas que originam as inflorescências, estas formadas nos ramos plagiotrópicos (Figuras 4A e 4B). Em geral, são encontradas quatro gemas nas axilas de ramos plagiotrópicos protegidas pela estípula interpeciolar e dispostas numa série de diferentes estádios de desenvolvimento como mostrado na Figura 4B, onde a gema mais desenvolvida encontra-se acima no ramo (G1) e a menos (G4) na base junto à conexão com o pecíolo.

O estabelecimento de um meristema axilar se dá inicialmente pelo crescimento do primeiro par de brácteas e coléteres em sua face adaxial (Figuras 4C e 4D). Após indução do estádio reprodutivo, os meristemas de inflorescência apresentam formação de um segundo par de brácteas, também revestido de coléteres em sua face adaxial, que secretam grande quantidade de mucilagem (Figuras 4E e 4F). Essa mucilagem irá acompanhar o desenvolvimento dos meristemas até o surgimento de todas as estruturas florais (APÊNDICE – Item 3). Nas amostras coletadas em maio, são observadas massas celulares adjuntas ao meristema central, indicando o início da formação dos meristemas florais (Figura 4F). E nas amostras coletadas em junho, observa-se uma clara diferença entre as fases de desenvolvimento, tanto dos meristemas de inflorescência, quanto dos meristemas florais (Figura 4G). O meristema floral central, mais desenvolvido, apresenta inicialmente o desenvolvimento de primórdios de sépala e pétala (Figura 4H) e, em seguida, formação de estames, carpelos e finalmente a formação do ovário (Figura 4I).



Figura 4 – Seções longitudinais mostrando a transição de um meristema vegetativo em meristema de inflorescência e a formação dos órgãos florais em *C. arabica*. Legendas: EI – estípulas interpeciolares, G1 a G4 – série de gemas axilares em diferentes estádios de desenvolvimento, Br – brácteas, Co – coléteres, Mu – mucilagem, MI – meristema da inflorescência, MF – meristema floral, S – sépala, P – pétala, E – estame, C – carpelo, Ov – ovário. Barras: figuras A, C, D e E = 100 µm; B, F, H e I = 500 µm; G = 1000 µm.

Observa-se nesse ponto o rompimento da barreira de mucilagem com o crescimento da estrutura floral (Ver APÊNDICE – Item 3).

Na figura 5 é apresentada em detalhe a formação dos primórdios dos órgãos florais. Inicialmente, o meristema floral adquire aspecto achatado por meio de divisões celulares da túnica e do corpo do promeristema na região periférica originada da zona central (Figura 5A). A partir da zona periférica são formados primeiramente os primórdios das sépalas e das pétalas (Figura 5B), seguidos dos primórdios dos estames e dos carpelos (Figura 5C e 5D). Observa-se também a presença de coléteres na lateral do corpo do meristema floral (Figura 5C), que estão presentes desde os estádios iniciais (Figura 5A). Esses coléteres são comumente observados nas bases de meristemas, tanto nos de inflorescência quanto nos florais e inicialmente recobrem todo o domo meristemático, adquirindo posição mediana com o posterior desenvolvimento das estruturas oriundas do meristema (Figuras 5A e C).

Há intensa atividade meristemática nas regiões centrais e periféricas do meristema floral, evidenciada pelo núcleo bastante denso das células (Figura 5E). A manutenção da protoderme ocorre por divisões anticlinais da camada L1 e há vacuolização dessas células em direção à região basal. A partir daí segue-se um período de expansão e crescimento dos órgãos florais terminando com a formação do ovário ínfero, bicarpelar, bilocular contendo os óvulos (Figura 5F). Um corte transversal de uma inflorescência apresenta dois botões florais já completamente formados com a arquitetura floral característica do cafeeiro: cinco pétalas, cinco estames e dois carpelos unidos, envoltos pelo segundo par de brácteas revestidas por coléteres que preenchem todo espaço entre as brácteas e os botões florais (Figura 5G). Os outros dois botões florais encontram-se em um nível de corte mais profundo (ver Figura 6I) devido às diferenças nos estádios de desenvolvimento dos mesmos.

24



Figura 5 – Figuras A a F, seções longitudinais do meristema floral mostrando a formação dos órgãos florais em *C. arabica.* Na figura G, um corte transversal da inflorescência com os botões florais já formados. Legendas: Tu – túnica (L1), Co – coléteres, S – sépala, P – pétala, E – estame, C – carpelo, Ov – ovário, O – óvulo, Pt – protoderme. Seta indica uma divisão anticlinal na L1. Barras: figuras A a F = 100 µm; G = 500 µm.

As análises morfológicas por microscopia eletrônica de varredura revelaram detalhes nas estruturas das gemas bem como dos órgãos florais formados posteriormente. Na série de imagens apresentada na Figura 6, observam-se as estípulas interpeciolares do caule que encobrem e protegem as gemas axilares no início do desenvolvimento (Figura 6A). Após indução do estado reprodutivo, ocorre o entumescimento das gemas, causado pelo crescimento e desenvolvimento do meristema de inflorescência, forçando a abertura das estípulas e permitindo seu surgimento (Figura 6B). Essa gema inicial apresenta seu primeiro par de brácteas abrangentes, cobrindo e fechando toda a estrutura. À medida que ocorre o desenvolvimento dos primórdios, as brácteas se afastam e os quatro botões florais da gema emergem (Figuras 6C e 6D). Em seguida, ocorre o crescimento dos botões florais que se projetam ainda mais para fora do corpo do meristema revelando o segundo par de brácteas opostas que os protegia internamente (Figura 6E). O arranjo dos órgãos florais antes da antese pode ser observado na figura 6F, na qual o estigma bifurcado parte dos estiletes unidos, o pistilo encontra-se cercado por duas das cindo anteras e por três das cinco pétalas. As sépalas reduzidas não aparecem na figura. Quando completado seu desenvolvimento, o estigma bipartido abre-se revelando suas papilas estigmáticas (Figuras 6G e 6H). Fica clara na estrutura floral completa a formação de duas estruturas acopladas, constituídas pelo corpo do meristema e um par de brácteas revestidas por coléteres, dispostas de forma cruzada onde se inserem os quatro botões florais (Figura 6I).

Essa caracterização morfo-anatômica serviu de base para o entendimento da formação das estruturas no tempo e espaço e posterior comparação com os cortes de hibridização *in situ*.



Figura 6 – Gemas axilares de *C. arabica* em microscopia eletrônica de varredura. Legendas: EI – estípulas interpeciolares, Br – bráctea, BF – botões florais, Co – coléteres, P – pétala, A - antera, Es – estigma, Et – estilete, Pe – papilas estigmáticas.

RT-PCR

Para facilitar o entendimento da grande quantidade de resultados, os genes MADS-box encontrados no cafeeiro foram citados pelo número correspondente à sua sequência, indicado também em todas as figuras.

As análises dos perfis de expressão por RT-PCR semi-quantitativa dos 23 prováveis genes MADS-box do cafeeiro selecionados revelaram que seis deles encontram-se expressos exclusivamente em órgãos vegetativos (1, 3, 3.1, 6, 17, 18) e 12 deles em órgãos reprodutivos ou em ambos (apenas órgãos reprodutivos: 9, 12, 14 e 21; em ambos: 2, 5, 10, 11, 13, 15, 16, 19), confirmando o que é relatado sobre a família quanto à sua importância durante o desenvolvimento reprodutivo (Figura 7). Apesar de amplificações terem sido observadas quando se utilizou os clones de cDNA como moldes para os genes 7 e 8, não foram observadas amplificações a partir das amostras de cDNA estudadas para os primers destes genes. Para três genes não foram feitas análises (4, 20 e 22), pois como citado anteriormente, suas sequências estavam incompletas. Os perfis de expressão de cada gene do cafeeiro puderam ser comparados com perfis de expressão relatados para genes homólogos.

Genes expressos no pool de tecidos reprodutivos da primeira análise tiveram suas expressões determinadas em cada fase do desenvolvimento floral e do fruto (respectivamente, Figuras 8A e 8B). As PCRs foram realizadas com dois números de ciclos diferentes, gerando reações saturadas ou não saturadas, servindo como repetição e revelando os padrões de expressão incipiente. Além de permitir a identificação das fases em que cada um desses genes são expressos, os experimentos de RT-PCR revelaram tendências de expressão ao logo do desenvolvimento reprodutivo.


Figura 7 – Perfis de expressão semi-quantitativa de genes MADS-box, obtidos por RT-PCR, em diferentes tecidos de *Coffea arabica*. Tecidos: R (raiz), C (caule), Fo (folha), A (ápice ou ponteiro), Fl (pool de estádios de flor) e Fr (pool de estádios de fruto); + (controle positivo, clone de cDNA como molde) e – (controle negativo, sem DNA molde). A tabela da direita indica onde foram consideradas expressões positivas.



Figura 8 – Perfis de expressão de genes MADS-box obtidos por RT-PCR semi-quantitativa: **A**) Em diferentes estádios do desenvolvimento floral em *C. arabica*. Tecidos: 1 – gemas de 2 mm de comprimento, 2 – gemas de 3 mm, 3 – gemas de 3 a 6 mm, 4 – gemas de 6 a 10 mm, 5 – gemas maiores que 10 mm e 6 – flor na antese. **B**) Em diferentes estádios do desenvolvimento dos frutos (antes da maturação) em *C. arabica*. Tecidos: 1' – frutos até 3 mm de comprimento ("chumbinho"), 2' – frutos de 3 a 4 mm, 3' – frutos de 4 a 5 mm, 4' – frutos de 5 a 10 mm, 5' – frutos de 10 a 15 mm e 6' – frutos maiores de 15 mm. Em ambas as figuras, + representa o controle positivo (clone de cDNA como molde) e -, o controle negativo (sem DNA molde).

A análise dos perfis de expressão dos genes de cafeeiro homólogos aos genes do modelo ABC, durante as fases do desenvolvimento floral, revelou que o gene *13-CaAP1* (função A) é expresso em todos os estádios, sendo a expressão mais acentuada no início do desenvolvimento (Figura 8A).

Genes da função B (5-CaAP3 e 10-CaPI) são fortemente expressos em estádios mais tardios. O gene 14-CaTM6 possui uma expressão menor, quando comparado com os dois genes anteriores, mas compartilha com eles a expressão mais acentuada nos dois últimos estádios.

O gene *12-CaAG* (função C) encontra-se expresso nos quatro últimos estádios com aumento da expressão nos dois últimos (Figura 8A).

A análise da expressão desses mesmos genes durante o desenvolvimento do fruto, mostra expressão semelhantes em todos os estádios para os genes *13-CaAP1* e *5-CaAP3*. Observou-se expressão durante estádios iniciais e finais para *14-CaTM6* (Figura 8B).

As quantidades de cDNAs em todas as amostras foram satisfatoriamente normalizadas pela intensidade das amplificações específicas do gene *GAPDH*. Uma vez considerado o nível de expressão deste gene como estável em todos os órgãos e tecidos analisados, a intensidade das bandas nos géis de agarose revelou de forma semi-quantitativa o nível de expressão dos genes MADS-box nas amostras consideradas. A forte intensidade das bandas dos controles positivos e a ausência de bandas nos controles negativos indicam que as RT-PCRs foram bem sucedidas, não apresentando contaminações. A amplificação de uma única banda, do tamanho previsto para cada par de primers, mostrou que os primers desenhados foram gene-específicos.

Hibridizações in situ

Todos os experimentos de hibridizações *in situ* dos ortólogos do modelo ABC em cafeeiro (*13-CaAP1*, *5-CaAP3*, *14-CaTM6*, *10-CaPI* e *12-CaAG*) produziram forte sinal de hibridização, possibilitando a estimativa de um padrão de expressão durante o desenvolvimento floral (Figuras 9, 10, 11, 12 e 13). As análises morfoanatômicas permitiram a identificação dos órgãos/tecidos onde os genes encontravam-se expressos. Para todos os genes foram obtidos sinais de expressão desde o início do desenvolvimento reprodutivo (meristemas de inflorescência) até a formação dos órgãos florais. A análise comparativa dos padrões de expressão para os cinco genes revelou que eles podem ser divididos em dois grupos principais complementares (ver Figura 14): o grupo dos genes que tem expressão em brácteas e tecido fundamental (córtex e medula) – *13-CaAPI*, *5-CaAP3* e *14-CaTM6*, e o grupo dos que tem expressão preferencialmente no promeristema e tecido vascular – *10-CaPI* e *12-CaAG*.

O gene *13-CaAP1* apresenta expressão inicial nas faces abaxial e adaxial dos dois pares de brácteas e no meristema fundamental periférico e central do meristema da inflorescência (Figura 9A). Essa expressão se mantém, sendo observada uma progressão da área hibridizada em direção aos primórdios de sépalas à medida que ocorre o crescimento e desenvolvimento do meristema floral (Figuras 9B e 9C). Esse padrão é bastante conservado e consistente durante os estádios de desenvolvimento dos órgãos florais, sendo estendido à face abaxial do primórdio da pétala à medida que este se desenvolve (Figuras 9D e 9E). Não se observou sinal de hibridização nos primórdios de estames e carpelos (Figura 9F). Os genes *5-CaAP3* e *14-CaTM6* apresentam padrão muito semelhante ao descrito para *13-CaAP1* (Figuras 10 e 11), porém com sinal menos intenso sugerindo possivelmente um menor nível de expressão.





Nos estádios iniciais, o gene *10-CaPI* apresenta expressão já no promeristema dos meristemas de inflorescência (Figura 12A). Este sinal de expressão permanece no promeristema dos meristemas florais, além de aparecer restrito ao procâmbio recém formado (Figura 12B). Posteriormente, o sinal de hibridização é detectado no início do desenvolvimento dos primórdios dos órgãos florais em camadas mais internas, sendo mantida a expressão no procâmbio (Figura 12C).



Figura 11 – Seções longitudinais do meristema floral de *C. arabica* mostrando a hibridização *in situ* do gene *14* – *CaTM6* (função B) em diferentes estádios do desenvolvimento floral. Barras: figuras A, C, D e F = 200 μ m, B e E = 500 μ m.

A expressão de *10-CaPI* é gradativamente diminuída nos primórdios de sépalas e pétalas (Figuras 12C, 12D e 12E), ficando assim restrita aos primórdios de estames e carpelos (Figura 12E e 12F). Adicionalmente, a figura 12G mostra um forte sinal de hibridização nos coléteres que revestem os meristemas florais.





Figura 12 – Seções longitudinais do meristema floral de *C. arabica* mostrando a hibridização *in situ* do gene 10 - CaPI (função B) em diferentes estádios do desenvolvimento floral. Barras: figuras A a F = 200 μ m, G = 500 μ m.

O gene *12-CaAG*, mostrou um padrão de expressão bastante semelhante ao observado para o gene *10-CaPI* (Figura 13), porém observou-se sinal de hibridização na face adaxial das anteras (Figura 13E) e em todo o primórdio do carpelo, principalmente na região adaxial do estilete e do ovário (Figura 13F). Da mesma forma que *10-CaPI*, esse gene apresentou expressão em coléteres (Figuras 13B e 13C).



Figura 13 – Seções longitudinais do meristema floral de *C. arabica* mostrando a hibridização *in situ* do gene 12 - CaAG (função C) em diferentes estádios do desenvolvimento floral. Barras: figuras A e F = 500 µm, B e D = 100 µm, C e E = 200 µm.



Figura 14 – Resumo esquemático das hibridizações *in situ* mostrando os locais onde foram considerados sinais positivos de expressão para os ortólogos dos genes ABC. Em parênteses a função relatada para cada um dos genes.

5. DISCUSSÃO

Características anatômicas

A sequência de desenvolvimento ontogenético dos verticilos florais em *C. arabica* é semelhante ao descrito para inúmeras outras espécies de angiospermas, diferenciando-se primeiramente as sépalas, em seguida as pétalas, os estames e finalmente os carpelos (Dedecca, 1957; Popham, 1963). A estrutura da inflorescência do cafeeiro reproduz a organização do sistema de ramificação vegetativa, sendo os dois pares de brácteas homólogos às folhas e estípulas interpeciolares dos ramos vegetativos (Meulen, 1939).

Meristemas de ramos ortotrópicos originam ramos plagiotrópicos ou ortotrópicos de substituição (Carvalho *et al.*, 1950) e por isso, podem ser considerados meristemas vegetativos que se mostram mais achatados e menores. Além disso, gemas vegetativas são consideradas indiferenciadas quando menores que 300 µm em *C. canephora* (Moens, 1963).

As gemas dos ramos plagiotrópicos ocorrem geralmente em número de quatro dispostas numa série de desenvolvimento (Wormer & Gituanja, 1970a), sendo G1 a mais desenvolvida. As gemas G1 podem formar ramos laterais ou inflorescências, enquanto as outras gemas raramente desenvolvem-se vegetativamente, podendo formar inflorescências ou permanecer indiferenciadas (Majerowicz & Söndahl, 2005).

Descrevendo as etapas do desenvolvimento floral, Majerowicz & Söndahl (2005) apontam que a passagem para o estado reprodutivo é marcada pelo entumescimento das gemas G1 e ligeira protuberância das estípulas. As observações histológicas desses autores apontam que ocorre alargamento e elevação do domo apical do meristema, que assume um aspecto tabular devido à intensa proliferação celular nas regiões laterais do ápice onde as células sofrem repetidas divisões anticlinais e periclinais. Nestas áreas, que originam o segundo par de brácteas, ocorre grande aumento do volume da gema com a presença de uma mucilagem ocreamarelada hidrofílica. Essa transição, ou seja, o período de indução reprodutiva ocorre aceleradamente entre janeiro e fevereiro, estendendo-se até julho. A data de coleta do material apresentado na figura 4F concorda com este período e estas observações, sendo considerada uma gema induzida contendo o meristema de inflorescência já em atividade reprodutiva. A mucilagem é provavelmente secretada pelos coléteres da região adaxial das brácteas (Figura 4 e APÊNDICE – item 3).

As brácteas são consideradas estruturas de suporte e proteção das gemas. Coletéres são estruturas multicelulares secretoras de mucilagem que lubrificam as gemas impedindo seu dessecamento e, em *C. arabica*, revestem as brácteas na face adaxial preenchendo todos os espaços ao redor dos meristemas florais. O fato do período de indução e início do desenvolvimento reprodutivo iniciar-se com a formação do segundo par de brácteas e intensa atividade de secreção dos coléteres, sugere que um mesmo mecanismo poderia estar ativando estes dois processos, conferindo proteção aos meristemas florais até o final de seu desenvolvimento (ver Figura 4 e APÊNDICE - item 3).

A partir da formação do segundo par de brácteas, opostas ao primeiro e algumas vezes não visíveis externamente (Majerowicz & Söndahl, 2005), segue-se a formação dos meristemas florais na região axilar, a partir do meristema de inflorescência. Sendo as brácteas homólogas às folhas e estípulas, alguns meristemas florais ectópicos podem ser eventualmente encontrados desenvolvendo-se na axila do primeiro par de brácteas. Os meristemas de inflorescência dão origem a quatro meristemas florais dispostos de forma cruzada. Os dois meristemas florais dispostos no eixo paralelo ao caule (ou alinhado com o primeiro par de brácteas) estão sempre num estádio mais desenvolvido que os meristemas florais do eixo perpendicular ao caule (ou alinhado com o segundo par de brácteas). A assincronia observada entre as gemas axilares e também entre os meristemas florais (ver figura 4G) poderia sugerir um mesmo mecanismo de controle no tempo de desenvolvimento. Porém, a assincronia entre as gemas axilares é estabelecida desde sua origem no caule e a desuniformidade do florescimento e da maturação são causadas principalmente pela assincronia no desenvolvimento dos meristemas florais oriundos das gemas G1 e G2, principais responsáveis pela produção das flores e frutos em *C. arabica* (Majerowicz & Söndahl, 2005). Portanto, os mecanismos atuantes no desenvolvimento das gemas axilares e dos meristemas florais estão separados no tempo e atuam provavelmente de maneira independente.

Os genes MADS-box em Coffea L.

A identificação de 23 genes MADS-box classificados como genes do tipo II foi confirmada pela análise de domínios conservados, com identificação da típica estrutura MIKC (exceto gene 7-*CaMd*). Genes do tipo I não foram encontrados, o que pode ser explicado pelo baixo nível de expressão ou por serem expressos sob condições não monitoradas em projetos transcriptoma (Dias *et al.*, 2005). Além disso, enquanto membros do grupo MIKC controlam efeitos fenotípicos óbvios e redundantes em mutantes de plantas-modelo, os dos grupos M α , M β e M γ teriam funções muito mais sutis, possivelmente não relacionadas ao desenvolvimento (Parenicova *et al.*, 2003).

Os perfis de expressão dos genes MADS-box em diferentes tecidos vegetativos e reprodutivos, confirmaram a importância destes genes para o para o desenvolvimento floral, assim como para o desenvolvimento de outras regiões vegetativas da planta e podem estar envolvidos em diversos processos, tais como, alongamento da raiz lateral, desenvolvimento de folhas, células-guarda e tricomas (Zhang & Forde, 1998; Alvarez-Buylla *et al.*, 2000b; Münster *et al.*, 2002b).

As análises comparativas de sequências indicaram a reconstrução dos diversos subgrupos dos homólogos de cafeeiro dos genes MADS-box, formando clados bem definidos (Becker & Theissen, 2003). Genes homólogos de uma mesma subfamília são, em geral, funcionalmente similares, podendo sugerir prováveis funções baseadas em ortologia (Kim *et al.*, 2005). Porém, tem-se demonstrado que apesar de serem originados de duplicação, podem sofrer modificações ao longo da evolução, desempenhando funções nem sempre conservadas (Kim *et al.*, 2005). A figura 3 demonstra a importância em ser considerada a distância evolutiva incluindo espécies próximas, pois dessa forma podem ser separados clados mais bem definidos e ortologias mais próximas. Duplicações recentes em algumas espécies tem resultado num grupo de parálogos que carregam partes das funções desempenhadas por seus homólogos (Soltis *et al.*, 2002). Esses prováveis eventos de subfuncionalização sugerem que enquanto as funções de membros individuais das subfamílias possam variar entre espécies, coletivamente estes genes teriam uma função conservada na regulação da identidade dos órgãos florais (Zahn *et al.*, 2005).

Para um melhor entendimento dessa relação entre os genes identificados em cafeeiro e seus homólogos, será feita uma discussão dentro de cada subfamília MIKC-like, dando-se maior ênfase àquelas correspondentes às funções A, B e C.

Subfamília SQUAMOSA-like (Função A)

São descritos quatro genes dessa subfamília em A. thaliana - APETALA1 (AP1), FRUITFULL (FUL), CAULIFLOWER (CAL) e AGL79 - sendo três deles caracterizados por mutantes. Os genes AP1 e CAL são parálogos parcialmente redundantes (Becker & Theissen, 2003), pois é o mutante duplo *ap1/cal* que produz o típico fenótipo *CAULIFLOWER* caracterizado pela extensiva proliferação de meristemas de inflorescência (Kempin *et al.*, 1995).

A expressão de *AP1* é primeiramente observada nos primórdios florais emergentes e, mais tarde, confinada às regiões mais periféricas do meristema floral, onde este gene está envolvido na especificação de sépalas e pétalas (Mandel *et al.*, 1992). Estudos recentes indicam que *AP1* atua como ponto central na rede regulatória envolvida na transição do meristema da inflorescência para o meristema floral (Kaufmann *et al.*, 2010). A participação deste gene na formação do perianto é recente, ocorrendo apenas em angiospermas superiores (Litt & Irish, 2003). O mutante *ful* apresenta defeitos na expansão e diferenciação de células das valvas nos frutos, dando ao fruto mutante um aspecto de "fruto cheio". Desta forma é considerado um gene determinante da identidade das valvas, que correspondem à maior parte da siliqua (Gu *et al.*, 1998).

Foram identificados em cafeeiro os genes 13 e 15, ortólogos aos genes *AP1* e *FUL*, respectivamente. Os perfis de expressão para *13-CaAP1* concordam com os dados publicados para outros homólogos de *AP1*, já que detectou-se expressão de *13-CaAP1* em ápices, botões florais e frutos. O gene *13-CaAP1* é preferencialmente expresso em tecidos reprodutivos e, em geral, a expressão se mantém consistentemente em todas as fases de desenvolvimento de flores e frutos.

A expressão de 15-CaFUL apresentou um padrão de expressão divergente do relatado para *A. thaliana* (Kim *et al.*, 2005). Detectou-se a expressão deste gene em tecidos de raiz e caule, e também em botões florais, não sendo detectados transcritos em amostras de fruto. Desta forma, espera-se que a função do gene 15-CaFUL seja diferente da função atribuída a seu possível ortólogo em Arabidopsis (FUL), já que nesta planta modelo, a principal função

atribuída a este gene tem relação com o desenvolvimento dos frutos. Isso pode ser explicado pelo fato da linhagem ancestral *FUL* que deu origem ao subclado *SQUA*, ter expressão mais ampla, incluindo órgãos vegetativos, como detectado em angiospermas basais (Kim *et al.*, 2005). Em contraste com o observado em *A. thaliana*, existe um maior número de parálogos de *FUL* em Solanaceae (Leseberg *et al.*, 2008). Segundo estes autores, este fenômeno favoreceu um aumento da complexidade fenotípica, sendo que interações diferenciais de proteínas produzidas por estes parálogos seriam as responsáveis pela divergência observada no desenvolvimento dos frutos das espécies em Solanaceae (Leseberg *et al.*, 2008).

Subfamílias DEF e GLO-like (Função B)

Diversos estudos demonstraram que as subfamílias DEF-like e GLO-like surgiram da duplicação de um gene ancestral na base do grupo das Angiospermas (Winter *et al.*, 2002), sendo por isso relatados como parálogos e agrupados em conjunto. O nome é dado ao grupo devido aos dois primeiros genes identificados: *DEFICIENS (DEF)* e *GLOBOSA (GLO)*, em *Antirrhinum* (Bowman *et al.*, 1989), sendo posteriormente descobertos seu respectivos ortólogos em *A. thaliana*, *APETALA3 (AP3)* e *PISTILLATA (PI)* (Sommer *et al.*, 1990; Tröbner *et al.*, 1992). Em ambas as espécies esses genes têm apresentado funções de formação de heterodímeros para estabelecimento da identidade das pétalas e estames (Riechmann *et al.*, 1996; Zachgo *et al.*, 1995).

Os genes pertencentes a essa classe são denominados genes B, com provável função ancestral de diferenciação dos órgãos reprodutivos masculinos de órgãos femininos (Winter *et al.*, 2002). Outra importante duplicação ocorrida na linhagem AP3, na base das eudicotiledôneas superiores, deu origem às linhagens euAP3 e TM6 (ou paleoAP3; Pnueli *et al.*, 1991; Kramer *et al.*, 1998). Essa duplicação teve grande importância na origem das

pétalas, claramente diferenciadas nesse grupo (Lamb & Irish, 2003). Dessa forma, sugere-se que genes euAP3, exclusivos de eudicotiledôneas superiores, estão associados à diferenciação de pétalas, enquanto que a função ancestral, de especificação de estames, foi mantida pela cópia TM6 (Vandenbussche *et al.*, 2004, de Martino *et al.*, 2006).

Foram encontrados três sequências deste grupo em cafeeiro, que puderam ser claramente diferenciada. Cada um destes genes, *10-CaPI*, *5-CaAP3* e *14-CaTM6*, mostraram-se preferencialmente expressos durante as fases finais do desenvolvimento de orgãos florais, sendo que *5-CaAP3* e *14-CaTM6* mostraram-se também expressos em diversas fases do desenvolvimento do fruto.

Esses genes por pertencerem ao modelo ABC, serão discutidos mais detalhadamente adiante.

Subfamília AG-like (Função C)

O gene *AGAMOUS* (*AG*) de *A. thaliana* foi o primeiro gene homeótico clonado dando nome ao grupo (Yanofsk *et al.*, 1990). Genes da subfamília *AG* são chamados de genes de função C por estarem envolvidos na especificação de estames e carpelos (Winter *et al.*, 1999). Homólogos de *AG* possuem padrões de expressão bastante conservados em Gimnospermas e Angiospermas basais (Jager *et al.*, 2003, Kim *et al.*, 2005), indicando uma função bastante antiga no grupo ancestral das plantas com sementes (Kramer *et al.*, 2003; Melzer & Theissen, 2007).

Outros três genes de *A. thaliana* são classificados como pertencentes à subfamília AG: SHATTERPROOF1 (SHP1), SHATTERPROOF2 (SHP2) e SEEDSTICK (SDK), sendo que AG e SHP1/2 exibem tanto funções redundantes quanto distintas (Pinyopich *et al.*, 2003). Enquanto AG é responsável por uma "bfunção C primaria", SHP1 e SHP2 são redundantes em promover identidade dos óvulos e carpelos, também desempenhando um papel na diferenciação do replum (Liljegren *et al.*, 2000). Em *Antirrhinum*, o gene *PLENA (PLE)* desempenha a função C (Carpenter & Coen 1990), enquanto o parálogo *FARINELLI (FAR)* é redundante e contribui para diferenciação dos estames (Davies *et al.*, 1999). Apesar da homologia funcional entre *AG* e *PLE*, estudos demonstraram que esses genes pertencem à linhagens parálogas, separadas após uma duplicação ocorrida antes da diversificação das eudicoteledoneas superiores (Davies *et al.*, 1999). Os múltiplos eventos de subfuncionalização que ocorreram nessa subfamília ilustram o potencial de duplicações gênicas na divergência de módulos genéticos, permitindo um aumento da diversidade morfológica (Kramer *et al.*, 2004).

Foram encontrados em cafeeiro dois genes pertencentes à subfamilia *AG*: *12-CaAG*, ortólogo a *FARINELLI* e proveniente da linhagem *AG* e o gene *22-CaSHP*, proveniente da linhagem *PLE* (Figura 3C). O perfil de expressão para o gene *12-CaAG* indicou que esse gene parece desempenhar a função C, já que é expresso em tecidos reprodutivos, durante algumas fases do desenvolvimento do fruto e nas fases finais do desenvolvimento floral, que coincide com a formação de estames e carpelos. A expressão nos verticilos internos foi confirmada pela hibridização *in situ*. O gene *22-CaSHP* não teve seu perfil de expressão determinado uma vez que a sequência deste gene revelou-se incompleta, mas um indício de sua expressão é o fato de que seus ESTs formadores são oriundos de bibliotecas obtidas a partir de botões florais.

Subfamília SEP (Função E)

O grupo dos genes *SEPALLATA (SEP)* é chamado de genes de função E, por serem essenciais à diferenciação dos quatro verticilos florais (Pelaz *et al.*, 2000). Proteínas SEP formam complexos com as proteínas das funções A, B e C, sendo suficientes para transformação de folhas em órgãos florais (Honma & Goto, 2001). Em *A. thaliana* são

encontrados quatro parálogos de *SEP*, sendo que somente com a obtenção de um mutante triplo (*sep1/sep2/sep3*) é observado um fenótipo claro, que apresenta a segundo e terceiro verticilios convertidos em sépalas e o desenvolvimento de uma nova inflorescência na região central (Pelaz *et al.*, 2000). A mutação *sep4* somada ao mutante triplo resulta na conversão de todos os órgãos florais em folhas, revelando uma importante função desse gene no desenvolvimento dos órgãos florais (Ditta *et al.*, 2004). Proteinas SEP atuam como importantes mediadores da formaçao de complexos se ordem superior entre os diferentes produtos dos genes MADS (Immink *et al.*, 2009).

Foram encontrados três genes do cafeeiro classificados nessa subfamília. O gene 8 apresentou-se ortólogo ao gene *SEP4*, e sua expressão não pôde ser determinada, provavelmente devido aos baixos níveis de expressão. O gene 11 apresentou homologia aos genes *SEP1* e *SEP2*, enquanto o gene 9 é provável ortólogo de *SEP3*. Os padrões de expressão dos genes 9 e 11 concordam com o relatado para *A. thaliana* (Flanagam & Ma, 1994). Immink *et al.*, (2009) sugerem que *SEP3* seja um ponto central na rede de interação de fatores de transcrição MADS-box, funcionando como um elemento comum na formação de complexos de ordem superior nos três verticilos internos da flor, o que coincide com os padrões de expressão do gene 9.

Subfamília AGL6-like

A subfamília *AGL6* está grupada num superclado junto com as subfamílias *SQUA* e *SEP*, porém a exata relação entre essas três linhagens permanece desconhecida (Becker & Theissen, 2003). Em *A. thaliana* são encontrados dois parálogos, *AGL6* e *AGL13*, para os quais não foram observados fenótipos nos mutantes, provavelmente devido à redundância desses genes e outros fatores (Koo *et al.*, 2010).

Enquanto o gene *AGL6* é expresso nos quatro verticilos florais (Mouradov *et al.*, 1998), o *AGL13* é restrito aos óvulos (Rounsley *et al.*, 1995). Recentemente foram relatados fenótipos em mutantes para estes genes em outras angiospermas, revelando que genes dessa classe estariam envolvidos na regulação das identidades dos órgãos florais e na determinação do meristema floral (Ohmori *et al.*, 2009). Diversos estudos indicaram que, embora os padrões de expressão em Gimnospermas e Angiospermas tenham sido mantidos nos órgãos masculinos e femininos (Yoo *et al.*, 2010), a expressão destes genes em órgãos masculinos não foi detectada em muitas linhagens (Viaen *et al.*, 2010).

No cafeeiro encontrou-se apenas o gene 21-CaAGL6 nessa subfamília, apresentando perfil característico de expressão em órgãos florais e em amostras de fruto. Embora não se tenha determinado em quais verticilos florais este gene seria expresso, um aumento da expressão no final do desenvolvimento floral sugere que 21-CaAGL6 possa estar associado à diferenciação de óvulos e carpelos.

Subfamília TM3-like

Membros da subfamília TM3 são encontrados tanto em Angiospermas quanto em Gimnospermas (Winter *et al.*, 1999) constituindo um clado muito antigo, com funções ancestrais no desenvolvimento vegetativo (Becker & Theissen, 2003). São encontrados seis representantes deste clado em *A. thaliana: SOC1 (AGL20), AGL14, AGL19, AGL42, AGL71 e AGL72*. Estes genes são preferencialmente expressos em partes vegetativas, mas também estão envolvidos na transição floral, sendo relatada a expressão em órgãos reprodutivos (Münster *et al.,* 2002a; Heuer *et al.,* 2001).

Foram encontrados em cafeeiro cinco homólogos pertencentes a esta classe, sendo três deles considerados parálogos muito próximos (3, 3.1 e 4). A formação de um clado

independente, mas sem o suporte de bootstrap, não permitiu relacioná-los aos grupos próximos (genes *AGL71/72* e *AGL14/19*). No entanto, os perfis de expressão indicaram estarem mais relacionados aos genes *AGL71* e *AGL72*, que mostram um amplo padrão de expressão, principalmente em tecidos vegetativos (Helliwell *et al.*, 2006), ao contrário dos genes *AGL14* e *AGL19* exclusivo de raízes (Rounsley *et al.*, 1995; Alvarez-Buylla *et al.*, 2000a). Os outros dois genes do cafeeiro (16 e 20) são respectivamente prováveis ortólogos dos genes *AGL42* e *SOC1*.

Homólogos ao gene *AGL42* são pouco estudados, apresentando expressão ampla tanto em tecidos vegetativos como reprodutivos (Helliwell *et al.*, 2006), o que concorda com a expressão observada em cafeeiro. Já o gene *SOC1* tem um mutante caracterizado, sendo descrito como um componente essencial da via pela qual o tempo de florescimento é regulado (Lee *et al.*, 2000) sendo também sugerido como formador de complexos protéicos superiores no desenvolvimento de partes vegetativas (Immink *et al.*, 2009). Uma vez que a sequência ortóloga a este gene em cafeeiro encontrava-se incompleta no clone correspondente, optou-se por não determinar seu perfil de expressão.

Subfamília AGL17-like

Os genes da classe *AGL17* são relatados como sendo preferencialmente expressos em raízes (Zhang & Forde, 1998; Burgeff *et al.*, 2002). Em *A. thaliana* são encontrados quatro genes pertencentes a esta classe: *AGL16*, *AGL17*, *AGL21* e *ANR1*. O único mutante identificado para um gene desta classe é o *anr1*, que não responde com a proliferação de raízes laterais em direção às zonas do solo ricas em nitrato, o que sugere uma função importante na cadeia de transdução de sinais pela qual as raízes são estimuladas pelo nitrato (Zhang & Forde, 1998).

Foram encontrados em cafeeiro dois genes parálogos, *1-CaAGL17-1* e *6- CaAGL17-2*, relacionados à classe de genes *AGL17*. O padrão de expressão encontrado em cafeeiro concorda com o observado para *AGL17* de *Arabidopsis*. Baixas expressões de *AGL17* podem ser encontradas também em tecido vegetativo (Han *et al.*, 2008), em concordância com a expressão de *1-CaAGL17-1* em tecidos de caule.

Subfamília SVP (STMADS11-like)

No cafeeiro foram encontrados três genes pertencentes a essa classe, 2-CaSVP-1, 17-CaSVP-2 e 18-CaAGL24, todos expressos em órgãos vegetativos. Isso concorda com estudos anteriores em Arabidopsis, em que transcritos do gene SVP são encontrados em folhas jovens e no meristema apical caulinar, sendo reprimidos no meristema da inflorescência por AP1 (Liu *et al.*, 2007). O mutante para o gene SVP, apresenta florescimento precoce, passando rapidamente pelos estágios vegetativos de desenvolvimento (Hartmann *et al.*, 2000). Esse gene codifica um repressor de florescimento, que prolonga todos os estágios vegetativos em plantas selvagens independentemente do fotoperíodo ou da vernalização (Becker & Theissen, 2003). Durante o desenvolvimento da flor, a expressão desse gene é fortemente reprimida (Hartmann *et al.*, 2000).

O gene *AGL24* é fortemente expresso no meristema apical e nos primórdios foliares. Durante a transição floral, a expressão é detectada também no meristema da inflorescência e depois no meristema floral, em estames e carpelos (Yu *et al.*, 2002). Curiosamente, transcritos do gene *18-CaAGL24*, provável ortólogo do gene *AGL24*, não foram observados em órgãos reprodutivos, talvez por ser difícil a detecção da expressão em fases iniciais durante a transição do meristema vegetativo para o reprodutivo, ou por essa função estar sendo representada pelo seu parálogo, *2-CaSVP-1*, que apresenta expressão em amostras de flores e frutos.

Subfamília FLC-like

Em A. *thaliana* o gene *FLC* (*Flowering Locus C*) é relatado como um inibidor de florescimento, já que seu mutante nulo apresenta florescimento precoce (Michaels & Amasino, 1999). O produto destes gene tem sido interpretado como um ponto de convergência para fatores ambientais e vias endógenas que regulam o tempo em que ocorrerá o florescimento (Becker & Theissen, 2003). O gene *FLC* é expresso em todas as partes das plantas, exceto em inflorescências, sendo regulado principalmente pela vernalização, mas também por vias autônomas e pela expressão de *FRI* (*FRIGIDA*; Sheldon *et al.*, 2000). Em cafeeiro, o provável ortólogo de FLC é o gene *19-CaFLC*, que possui um padrão de expressão similar ao seu representante em *Arabidopsis*, ocorrendo em todos os tecidos e bem pouco expresso em botões florais, conforme os dados obtidos nos experimentos de RT-PCR.

Genes da subfamília *FLC* de muitas espécies respondem à vernalização, sugerindo que esses genes constituem um mecanismo evolutivo de adaptação às condições temperadas, pois, assim como genes *AGL15*, não são encontrados em representantes fora da família *Brassicaceae* (Becker & Theissen, 2003). Esses genes poderiam ser, portanto, indicadores do término e início de uma época mais favorável para desenvolvimento e propagação.

O modelo ABC em C. arabica L.

O modelo ABC postula que três funções (A, B e C) atuam de forma combinatória para conferir a identidade dos órgãos florais em cada verticilo (Coen & Meyerowitz, 1991). Esse programa básico é amplamente conservado entre as espécies com flores (Krizek & Fletcher, 2005), apesar da possibilidade de ocorrência de diferenças caracterizadas por eventos de duplicação e subsequente divergência gênica entre espécies (Litt & Irish, 2003).

No cafeeiro, as hibridizações *in situ* indicaram um perfil de expressão aparentemente bastante diferente dos relatados para *Arabidopsis*, mostrando que os ortólogos dos genes *AP1* (*13-CaAP1* - função A), *AP3 e TM6* (*5-CaAP3* e *14-CaTM6* - função B) possuem um padrão de expressão similar, com transcritos detectados nas partes estéreis da flor (brácteas, corpo do meristema e perianto), enquanto que os ortólogos do gene *PI* (*10-CaPI* – função B) e *AG* (função C), são expressos nos órgãos férteis.

A função A conferida pelo gene *AP1* em *A. thaliana* determina a especificação da identidade do meristema floral bem como a diferenciação do perianto (Weigel & Meyerowitz, 1994). No cafeeiro, as hibridizações *in situ* confirmaram o padrão de expressão de *13-CaAP1* em estádios iniciais, importantes para a determinação da identidade do meristema floral e para o desenvolvimento de brácteas e sépalas (Figuras 9A, 9B e 9C). No entanto, o gene não apresentou expressão nas fases iniciais de desenvolvimento das pétalas (Figura 9D) e, apesar de serem relatadas expressões mais amplas de homólogos em outras espécies (Kim *et al.* 2005), outro fato parece explicar essa expressão.

Em *Arabidopsis*, os genes que desempenham as funções B e C basicamente diferenciam órgãos florais férteis de estéreis: carpelos pela expressão apenas de genes C no quarto verticilo e estames pela expressão simultânea de genes das classes B e C no terceiro verticilo (Coen & Meyerowitz, 1991). Dessa forma, no início do desenvolvimento dos meristemas florais, a expressão dos genes *10-CaPI* e *12-CaAG* são expressos nas células da região de origem dos órgãos férteis (Figuras 12C, 12D e 13C). As identidades de pétala e estame são discriminadas apenas tardiamente, com o acréscimo dos transcritos de *13-CaAP1* nos primórdios de pétalas, juntamente com *5-CaAP3* e *14-CaTM6*. As linhagens *AP3/PI* e *AG* presentes em gimnospermas revelam que a origem de genes da classe B e C são anteriores à origem das flores, sugerindo uma função primordial na especificação de órgãos reprodutivos

anterior à formação do perianto (Winter *et al.*, 1999). Assim, ocorre primeiramente a diferenciação da região que origina os estames e posteriormente a determinação das pétalas pelo gradativo aumento de expressão na face abaxial de *13*-CaAP1, 5-*CaAP3* e *14*-*CaTM6* (ver Figura 14).

Desta forma, a expressão dos ortólogos dos genes das classes B e C (10-CaPI e 12-CaAG) passam a exercer as funções conservadas na especificação dos órgãos reprodutivos que originam os estames e carpelos (Kim *et al.*, 2005). O gene 12-CaAG mantém sua expressão até a formação completa do carpelo e se restringe à região das anteras nos estames, onde provavelmente serão necessários para indução da microsporogênese pela ativação de homólogos do gene SPOROCITLESS (Ito *et al.*, 2004).

Os ortólogos dos genes da classe B (5-CaAP3, 14-CaTM6 e 10-CaPI), foram observados em todos os verticilos durante todas as fases do estabelecimento dos órgãos florais. Genes da classe B atuam basicamente na diferenciação do perianto e dos órgãos masculinos de femininos (Kramer & Irish 1999), sendo ambos os genes AP3 e PI, em Angiospermas, requeridos para o desenvolvimento apropriado do órgão floral. Mutantes de *A. thaliana* para esses genes resultam em mudanças homeóticas com perda da identidade do estame e em *A. majus*, conversão de estames em estruturas femininas (Jack *et al.*, 1992; Tröbner *et al.*, 1992). Com a duplicação da linhagem AP3, resultando na formação dos parálogos *euAP3* e *TM6* (Kramer *et al.*, 1998), estudos têm demonstrado diferentes funções para o gene *euAP3*, atuando no desenvolvimento mais específico da pétala, mas contribuindo para diferenciação de estames (de Martino *et al.*, 2006; Rijpkema *et al.*, 2006). Dessa forma, genes ortólogos da classe B teriam adquirido distintos papéis funcionais, já que o complexo multiprotéico da classe B como um todo teria retido uma função comum (Geuten & Irish, 2010). No cafeeiro, os ortólogos dos

genes *AP3* e *TM6* parecem envolvidos em especificar pétalas (ver Figura 14). O fato de não estarem expressos em estames e carpelos indica que não estão envolvidos no desenvolvimento desses órgãos, cuja identidade é provavelmente determinada pelo ortólogo ao gene *PI*, o gene *10-CaPI* em *C. arabica*.

Disso surgem duas importantes questões: como são explicadas as expressões não esperadas para eudicoltiledoneas superiores de ortólogos B (5-CaAP3 e 14-CaTM6) em sépalas e em carpelos (10-CaPI)? Além disso, sendo 13-CaAP1 e CaAP3/CaTM6 expressos em sépalas e pétalas, bem como 10-CaPI e 12-CaAG em estames e carpelos, como podem se diferenciar esses quatro verticilos? Algumas hipóteses discutidas a seguir poderiam ajudar a explicar essas questões.

Para responder a primeira pergunta, temos que padrões amplos de genes da classe B são encontrados em diversas angiospermas basais, podendo ter sido mantido em algumas espécies ao longo da evolução (Kim *et al.*, 2005). Porém, parece mais provável que a expressão desses genes em sépalas, assim como em outras regiões, esteja relacionada à formação de complexos com *13-CaAP1* e *5-CaAP3* contribuindo para alongamento e proliferação celular sem especificação da identidade do órgão, numa possível rota de ativação de giberelinas orquestrada pela expressão *de AP1* (Kaufmann *et al.*, 2010).

Além disso, uma análise da expressão de genes do grupo E durante o desenvolvimento floral poderia revelar a formação de complexos específicos necessários apenas para a identidade de determinado verticilo, mostrando que apesar de ocorrer a expressão de outros genes e da aparente redundância de genes *SEP*, eles atuem de forma específica na formação de complexos em determinadas regiões (Immink *et al.*, 2009). Pode-se especular também, que a expressão em tecidos vasculares de *10-CaPI* contribui de alguma forma para o

54

desenvolvimento da pétala, o que não ocorre para sépalas (ver Figura 12E). Funções adicionais combinadas no câmbio têm sido recentemente descritas para algumas proteínas MADS-box (Melzer *et al.*, 2009). Para o fato de *10-CaPI* estar expresso em carpelos, os resultados das RT-PCR semi-quantitativas mostram que a expressão tende a diminuir ao longo do desenvolvimento floral, mas infelizmente não pode ser confirmada a região em que estava restrita à expressão desse gene na última fase de desenvolvimento.

Para a segunda pergunta, Theissen & Melzer (2007) propoem em uma revisão que a transição no modo de interação protéica de dímeros para tetrâmeros teria aprimorado os mecanismos gênicos regulatórios pelo aumento da cooperatividade entre proteínas MIKC-like que se ligam ao DNA. Isso quer dizer que, no caso de ligações cooperativas entre fatores de transcrição, um mecanismo de alteração genética pode ser criado por pequenas mudanças na concentração da proteína causando uma resposta dramática no gene alvo regulado podendo resultar em mudanças drásticas nos programas de desenvolvimento (Ptashne, 2004; 2005).

De acordo com isso, é proposto um gráfico baseado na intensidade de expressão das RT-PCRs semi-quantitativas para os genes 13-CaAP1, 5-CaAP3, 14-CaTM6, 10-CaPI e 12-CaAG, no qual a especificação dos órgãos florais são explicados pelas diferenças de concentração desses genes em cada etapa do desenvolvimento floral (ver APÊNDICE – Item 4). Para facilitar o entendimento, o gráfico foi dividido em duas partes mostrando separadamente os dois grupos de expressão similares (grupo CaAP1/CaAP3/CaTM6 - Gráfico A, e grupo CaP1/CaAG – Gráfico B) ao longo do desenvolvimento floral. No primeiro gráfico, as primeiras fases de desenvolvimento, onde são especificadas as sépalas, tem a expressão do gene 13-CaAP1 (Função A) maior que a de 5-CaAP3 (Função B), mostrando uma maior atuação da função A nesse período. Nas etapas seguintes, que coincidem com as etapas de desenvolvimento das pétalas (ver resultados análises morfo-anatômicas), ocorre aumento da expressão do gene 5-CaAP3 que ultrapassa o nível da expressão do gene 13-CaAP1, evidenciando uma maior atuação da função B, auxiliada também pela da expressão do gene 14-CaTM6 (ver APÊNDICE – Item 4 Gráfico A). Dando embasamento a esses resultados, Kaufmann *et al.* (2010) sugerem que *AP1* controla de maneira complexa um módulo regulatório de genes que asseguram e auto-regulam sua própria expressão, suprimindo repressores florais e genes de identidade de meristema em primórdios florais emergentes.

No segundo gráfico é observado que as expressões dos genes *10-CaPI* e *12-CaAG* é menor nas primeiras etapas do desenvolvimento, sendo aumentadas a partir do estádio 2, onde são diferenciados os órgãos férteis (estames e carpelos) dos não férteis do perianto. O gene *10-CaPI* (função B) tem expressão bem superior ao gene *12-CaAG* (função C) nas etapas 3 e 4, quando se formam os estames, ao passo que nas etapas finais onde se estabelecem os carpelos, o gene *12-CaAG* tem sua expressão bastante aumentada demonstrando uma maior atuação da função C (ver APÊNDICE – Item 4 Gráfico B).

Embora tenha sido relatada análises quantitativas por qRT-PCR para alguns desses genes durante o desenvolvimento reprodutivo em *C. arabica* (Oliveira *et al.*, 2010), os dados são escassos para as diversas fases do desenvolvimento floral e não incluem todos os ortólogos do modelo ABC. Sendo assim, não permitem confirmar as hipóteses do estabelecimento da identidade dos órgãos florais baseadas nos níveis de expressão, deixando em aberto a necessidade de uma análise quantitativa mais exata.

Dessa forma, é sugerido que os genes do modelo ABC no cafeeiro, embora expressos de maneira mais ampla do que os relatados em eudicotiledôneas superiores, atuem de maneira conservada sendo principalmente regulados pela concentração de seus transcritos em cada verticilo floral em desenvolvimento. Lembrando que *C. arabica* é uma espécie relativamente

recente (Lashermes *et al.*, 1999), genes MADS-box de identidade floral poderiam ter adquirido novas funções combinatórias resultando em mecanismos sofisticados de desenvolvimento, como os apresentados acima, e não só restritos aos órgãos florais, como por exemplo, ativando o início do desenvolvimento dos meristemas florais juntamente com a ativação da secreção de mucilagem pelos coléteres.

6. CONCLUSÕES

A partir do banco CAFEST foram obtidas 23 sequências MADS-box em *C. arabica*, sendo 20 completas. O elevado número de reads e o baixo nível de redundância nas diferentes bibliotecas permitiram visualizar diferenças alélicas e splicings alternativos de cada sequência. As análises filogenéticas classificaram as 23 sequências em suas respectivas subfamílias e indicaram os ortológos do modelo ABC. Os locais de expressão RT-PCR indicaram de forma semi-quantitativa os tecidos onde os genes estariam sendo transcritos no cafeeiro, bem como durante as fases do desenvolvimento reprodutivo. Dessa forma, todas as sequências puderam ser comparadas aos respectivos homólogos em *A. thaliana*, sugerindo prováveis funções conservadas. Com as análises morfo-anatômicas foi determinada a série de desenvolvimento de um meristema para estabelecimento dos órgãos florais e a hibridização *in situ* identificou em quais desses locais os ortólogos do modelo ABC estariam sendo requisitados ao longo do desenvolvimento.

Foi constatado que genes do modelo ABC atuam de maneira diferenciada no tempo e espaço durante o desenvolvimento dos verticilos florais quando comparados a outras espécies, porém, baseado em seus níveis de expressão, a maneira combinatória de atuação de cada um dos fatores permanece relativamente conservada. Resumindo, genes ortólogos das funções B (10-CaPI) e C (12-CaAG) determinam os locais onde vão se desenvolver os órgãos reprodutivos, o gene ortólogo da função A (*13-CaAP1*) e B (*5-CaAP3* e *14-CaTM6*) determinam primeiramente as sépalas e posteriormente tem sua expressão estendidas às pétalas culminando, por isso, na formação de um estame epipétalo. Os níveis dos genes ortólogos de cada um dos fatores determinam a diferenciação entre verticilos que possuem expressões similares.

Sugere-se que genes ortólogos do modelo ABC tenham obtido funções diferenciais e específicas durante a evolução, como por exemplo, a ativação de um mecanismo de proteção por secreção dos coléteres ou desenvolvimento de tecido vascular. Essa observação é pertinente, tendo-se em vista que eventos de duplicação seguidos de funcionalização ao longo do tempo são relatados como um dos principais mecanismos de ganho de função (Theissen & Melzer, 2007) e, para o cafeeiro em particular, a estrutura de seu genoma tem forte potencial para isso [lembrando se tratar de um alotetraplóide recente com genomas de duas diferentes espécies diplóides e de fecundação cruzada (Lashermes *et al.*, 1999)]. Concluindo, estes genes parecem ter adquirido funções específicas ao longo do desenvolvimento do cafeeiro, mas conservando de maneira robusta mecanismos reguladores essenciais observados também em outras espécies.

Dessa forma, o trabalho alcançou seus objetivos com a identificação e análise dos ortólogos do modelo ABC durante o desenvolvimento reprodutivo do cafeeiro e, além disso, abriu enormes possibilidades de estudo para diversos processos que envolvem outros genes MADS-box também identificados. *Coffea arabica* é uma espécie com características peculiares que por isso tem muito a oferecer na elucidação de processos de florescimento, bem como, com o melhoramento dos métodos de produção no incremento da qualidade da bebida, já há tempos apreciada em todo o mundo.

58

7. REFERÊNCIAS BIBLIOGRÁFICAS

- Altschul, S.F.; Madden, T.L.; Schaffer, A.A.; Zhang, J.; Zhang, Z.; Miller, W. & Lipman, D.J. (1997) Gapped BLAST and PSI BLAST: A new generation of protein database search programs. Nucleic Acids Research, 25: 3389-3402.
- Alvarez-Buylla, E.R.; Pelaz, S.; Liljegren, S.J.; Gold, S.E.; et al. (2000a) An ancestral MADS-box gene duplication occurred before the divergence of plants and animals. Proc. Natl Acad. Sci. USA, 97: 5328-5333.
- Alvarez-Buylla, E.R.; Liljegren, S.J.; Pelaz, S.; Gold, S.E.; Burgeff, C.; Ditta, G.S.; Vergara-Silva, F. & Yanofsk, M.F. (2000b) MADS-box gene evolution beyond flower: expression in pollen, endosperm, grand cells, roots and trichomes. The Plant J., 24: 457-466.
- Bailey, T.L. & Elkan, C. (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. In Proceeding of the Second International Conference on Intelligent Systems for Molecular Biology (Menlo Park, CA: AAAI Press), pp. 28–36.
- Barsalobres-Cavallari, C.F.; Severino, F.E.; Maluf, M.P. & Maia, I.G. (2009) Identification of suitable internal control genes for expression studies in *Coffea arabica* under different experimental conditions. BMC Mol. Biol., 10:1-11.
- **Becker, A. & Theißen, G.** (2003) The major clades of MADS-box genes and their role in the development and evolution of flowering plants. Molecular Phylogenetics and Evolution, 29: 464-489.
- Benlloch R, Berbel A, Serrano-Mislata A, Madueno F. (2007) Floral initiation and inflorescence architecture: a comparative view. Annals of Botany, 100 (3): 659–676.
- Bowman, J.L.; Smyth, D.R. & Meyerowitz, E.M. (1989) Genes directing flower development in Arabidopsis. The Plant Cell, 1: 37-52.
- Burgeff, C.; Liljegren, S.J.; Tapia-López, R.; Yanofsky, M.F.; Alvarez-Buylla, E.R. (2002) MADS-box gene expression in lateral primordia, meristems and differentiated tissues of *Arabidopsis thaliana* roots. Planta, 214: 365-372.
- **Camargo, A.P. & Camargo, M.B.P.** (2001) Definição e esquematização das fases fenológicas do Cafeeiro Arábica nas condições tropicais do Brasil. Bragantia, 60 (1): 65-68.
- **Camargo, A.P.** (1985) Florescimento e frutificação do café arábica nas diferentes regiões cafeeiras do Brasil. Pesq. Agrop. Bras., 20 (7): 831-839.
- Carpenter, R. & Coen, E.S. (1990) Floral homeotic mutations produced by transposon-mutagenesis in *Antirrhinum majus*. Genes Development, 4: 1483–1493.
- Carvalho, A.; Krug, C.A.; Mendes, J.E.T. (1950) O dimorfismo dos ramos em *Coffea arabica* L. Bragantia, 10 (6): 151-159.
- **Charrier, A. & Berthaud, J.** (1985) Botanical classification of coffee. In Clifford MN, Wilson KC (Eds), Coffee: botany, biochemistry and production of beans and beverage. Croom Helm, London, pp. 13-47.
- Coen, E.S. & Meyerowitz, E.M. (1991) The war of the whorls: genetic interactions controlling flower development. Nature, 353: 31-37.
- Davies, B.; Motte, P.; Keck, E.; Saedler, H.; Sommer, H.; Schwarz-Sommer, Z. (1999) PLENA and FARINELLI: redundancy and regulatory interactions between two *Antirrhinum* MADS-box factors controlling flower development. EMBO J., 18: 4023–4034.

- Davis, A.P.; Govaerts, R.; Bridson, D.M.; Stoffelen, P. (2006) An annotated checklist of the genus *Coffea* L. (Rubiaceae). Botanical J. of the Linnean Society, 152: 465–512.
- De Bodt, S.; Raes, J.; Florquin, K.; Rombauts, S.; Rouze, P.; Theissen, G.; Van de Peer Y. (2003) Genomewide structural annotation and evolutionary analysis of the type I MADS-box genes in plants. J. Mol. Evol. 56: 573-586.
- de Martino, G.; Pan, I.; Emmanuel, E.; Levy, A. & Irish, V.F. (2006). Functional analyses of two tomato apetala3 genes demonstrate diversification in their roles in regulating floral development. The Plant Cell 18: 1833–1845.
- **Dedecca, D.M.** (1957) Anatomia e desenvolvimento ontogenético de *Coffea arabica* L. var. Typica Cramer. Bragantia 16 (23): 315-366.
- Dias, B.F.O.; Simões-Araújo, J.L.; Russo, C.A.M.; Margis, R. & Alves-Ferreira, M. (2005) Unravelling MADS-box gene family in Eucalyptus spp.: A starting point to na understanding of their developmental role in trees. Genetics and Molecular Biology, 28 (3, suppl): 501-510.
- Ditta, G.; Pinyopich, A.; Robles, P.; Pelaz, S.; Yanofsky, M.F. (2004) The SEP4 gene of Arabidopsis thaliana functions in floral organ and meristem identity. Current Biology 14: 1935–1940.
- Fasio, L.H. & Silva, A.E.S. (2007) Importância econômica e social do café Conilon. In: FERRÃO, R. G. et al. (Ed.). Café Conilon. Vitória, ES: Incaper, cap. 1, p. 37-52.
- Flanagan, C.A. & Ma, H. (1994) Spatially and temporally regulated expression of the MADS-box gene AGL2 in wild-type and mutant Arabidopsis flowers. Plant Mol. Biol., 26: 581-595.
- Frederico, D. & Maestri, M. (1970) Ciclo de crescimento dos botões florais de café (*Coffea arabica* L.). Ver. Ceres, Viçosa, 17 (92): 171-181.
- Geuten, K. & Irish, V. (2010). Hidden Variability of Floral Homeotic B Genes in Solanaceae Provides a Molecular Basis for the Evolution of Novel Functions. The Plant Cell 22: 2562-2578.
- **Gouveia, N.M.** (1984) Estudo da diferenciação e crescimento das gemas florais de *Coffea arabica* L.: observações sobre antese e maturação dos frutos. Univ. Estadual de Campinas, Dissertação (Mestrado).
- **Gramzow, L. & Theissen, G.** (2010) A hitchhiker's guide to the MADS world of plants. Genome Biology, 11: 214-225.
- Gramzow, L.; Ritz, M.S.; Theissen, G. (2010) On the origin of MADS-domain transcription factors. Trends Genetics, 26: 149-153.
- Grassias, M. & Kammacher, P. (1975) Observation sur la conjugaison chromosomique de *Coffea arabica* L. Café Cacao Thé 19: 177-190.
- Gu, Q.; Férrandiz, C.; Yanofsky, M. F.; Martienssen, R. (1998) The FRUITFULL MADS-box gene mediates cell diferentiation during Arabidopsis fruit development. Development, 125: 1509-1517.
- Han, P.; Garcia-Ponce, B.; Fonseca-Salazar, G.; Alvarez-Buylla, E.R.; Yu, H. (2008) AGAMOUS-LIKE 17, a novel flowering promoter, acts in a FT-independent photoperiod pathway. The Plant J., 55: 253-265.
- Harlan, J.R.; deWet, J.M.J. (1975) On Ö. Winge and a prayer: the origins of polyploidy. Bot. Rev., 41: 361-190.
- Hartmann, U.; Höhmann, S.; Nettesheim, K.; Wisman, E.; Saedler, H.; Huijser, P. (2000) Molecular cloning of SVP: a negative regulator of the floral transition in Arabidopsis. The Plant J., 21: 351–360.

- Helliwell, C.A.; Wood, C.C.; Robertson, M.; James Peacock, W.; Dennis, E.S. (2006) The Arabidopsis FLC protein interacts directly in vivo with SOC1 and FT chromatin and is part of a high-molecular-weight protein complex. The Plant J., 46: 183-192.
- Henikoff, S. & Henikoff, J.G. (1997) Embedding strategies for effective use of information from multiple sequence alignments. Protein Science, 6: 698-705.
- Heuer, S.; Hansen, S.; Bantin, J.; Brettschneider, R.; Kranz, E.; Lörz, H.; Dresselhaus, T. (2001) The maize MADS box gene ZmMADS3 affects node number and spikelet development and is co-expressed with ZmMADS1 during flower development, in egg cells, and early embryogenesis. Plant Phys., 127: 33-45.
- Honma, T. & Goto, K. (2001) Complexes of MADS-box proteins are sufficient to convert leaves into floral organs. Nature, 409: 525-529.
- Huang, X. & Madan, A. (1999) CAP3: A DNA sequence assembly program. Genome Research, 9: 868-877.
- Immink, R.G.H.; Tonaco, I.A.N.; Folter, S.; Shchennikova, A.; Dijk, A.D.J.; Busscher-Lange, J.; Borst, J.W.; Angenent, G.C. (2009) SEPALLATA3: the "glue" for MADS box transcription factor complex formation. Genome biology 10: R24.
- Ito, T.; Wellmer, F.; Yu, H.; Das, P.; Ito, N.; Alves-Ferreira, M.; Riechmann, J.L.; Meyerowitz, E.M. (2004) The homeotic protein AGAMOUS controls microsporogenesis by regulation of *SPOROCYTELESS*. Nature, 430: 356–360.
- Izawa, T.; Takahashi, Y. & Yano, M. (2003) Comparative biology comes into bloom: genomic and genetic comparison of flowering pathways in rice and *Arabidopsis*. Curr. Opin. Plant Biol., 6: 113-120.
- Jack, T.; Brockman, L. & Meyerowitz, E.M. (1992). The homeotic gene APETALA3 of Arabidopsis thaliana encodes a MADS box and is expressed in petals and stamens. Cell, 68: 683–697.
- Jager, M.; Hassanin, A.; Manuel, M.; Le Guyader, H. & Deutsch, J. (2003) MADS-box genes in Ginkgo biloba and the evolution of the AGAMOUS family. Mol. Biol. Evol., 20: 842–854.
- Kaufmann, K.; Wellmer, F.; Muiño, J.M.; et al. (2010) Orchestration of floral initiation by APETALA1. Science, 328: 85-89.
- Kempin, S.A.; Savidge, B.; Yanofsky, M.F. (1995) Molecular basis of the cauliflower phenotype in *Arabidopsis*. Science, 267: 522–525.
- Kim, S.; Koh, J.; Yoo, M-J.; Kong, H.; Hu, Y.; Ma, H.; Soltis, P.S.; Soltis, D.E. (2005) Expression of floral MADS-box genes in basal angiosperms: implications for the evolution of floral regulators. The Plant J., 43: 724-744.
- Koo, S.C.; Bracko, O.; Park, M.S.; Schwab, R.; Chun, H.J.; et al. (2010) Control of lateral organ development and flowering time by the Arabidopsis thaliana MADS-box gene AGAMOUS-LIKE6. The Plant J., 62 (5): 807-816.
- Kramer, E.M. & Irish, V.F. (1999). Evolution of genetic mechanisms controlling petal development. Nature, 399: 144–148.
- Kramer, E.M.; Dorit, R.L.; Irish, V.F. (1998) Molecular evolution of genes controlling petal and stamen development: duplication and divergence within the *APETALA3* and *PISTILLATA* MADS-box gene lineages. Genetics. 149:765–783.
- Kramer, E.M.; Di Stilio, V.S.; & Schluter, P.M. (2003) Complex patterns of gene duplications in the *APETALA3* and *PISTILLATA* lineages of the Ranunculaceae. Int. J. Plant Sci. 164: 1-11.

- Kramer, E.M.; Jaramillo, M.A.; Di Stilio, V.S. (2004) Patterns of gene duplication and functional evolution during the diversification of the AGAMOUS subfamily of MADS box genes in angiosperms. Genetics. 166: 1011–1023.
- Krizek, B.A. & Fletcher, J.C. (2005) Molecular mechanisms of flower development: an armchair guide. Nature, 6: 688-698.
- Lamb, R.S. & Irish, V. (2003) Functional divergence within the APETALA3/PISTILLATA floral homeotic gene lineages. Proc. Natl. Acad. Sci. USA., 100 (11): 6558-6563.
- Letunic, I.; Goodstadt, L.; Dickens, N. J.; Doerks, T.; *et al.* (2002) Recent improvements to the SMART domain-based sequence annotation resource. Nucleic Acids Research, 30: 242-244.
- Liljegren, S.J.; Ditta, G.S.; Eshed, Y.; Savidge, B.; Bowman, J.L.; Yanofsky, M.F. (2000) SHATTERPROOF MADS-box genes control seed dispersal in *Arabidopsis*. Nature, 404: 766-770.
- Litt, A. & Irish, V.F. (2003) Duplication and Diversification in the *APETALA1/FRUITFULL* Floral Homeotic Gene Lineage: Implications for the Evolution of Floral Development. Genetics, 165: 821-833.
- Litt, A. & Kramer, E.M. (2010) The ABC model and the diversification of floral organ identity. Sem. in Cell & Dev. Biol., 21: 129-137
- Liu, C.; Zhou, J.; Bracha Drori, K.; Yalovsky, S.; Ito, T.; Yu, H. (2007) Specification of Arabidopsis floral meristem identity by repression of flowering time genes. Development, 134: 1901-1910.
- Majerovicz, N.; Söndahl M.R. (2005) Induction and differentiation of reproductive buds in *Coffea arabica* L. Braz. J. Plant Physiol., 17 (2): 247-254.
- Mandel, M.A.; Gustafson-Brown, C.; Savidge, B.; Yanofsky, M.F. (1992) Molecular characterization of the *Arabidopsis* floral homeotic gene *APETALA1*. Nature, 360: 273-277.
- Maurin, O.; Davis, A. P.; Chester, M.; Mvungi, E. F.; Jaufeerally-Fakim, Y. & Fay, M F. (2007) Towards a Phylogeny for *Coffea* (Rubiaceae): identifying well-supported lineages based on nuclear and plastid DNA sequences. Annals of botany, 100 (7): 1565-1583.
- Melzer, S.; Lens, F.; Gennen, J.; Vanneste, S.; Rohde, A.; Beeckman, T. (2009) Flowering-time genes modulate meristem determinacy and growth form in *Arabidopsis thaliana*. Nat. Genet. 40: 1489-1492.
- Messenguy, F. & Dubois, E. (2003) Role of MADS box proteins and their cofactors in combinatorial control of gene expression and cell development. Gene, 316:1-21.
- Meulen, A. van der (1939) On the structure and periodical development of the flower buds of coffee species. Proc. Roy. Neth. Acad. of Sci. 38 (2): 1-128.
- Michaels, S.D. & Amasino, R.M. (1999) *Flowering Lócus C* encodes a novel MADS domain protein that act as a repressor of flowering. The Plant Cell, 11: 949-956.
- Moens, P. (1963) Les bourgeons végétatifs et génératifs de *Coffea* canephora Pierre. (Étude morphologique et morfogénétique). La cellule, 63 (2): 165-244.
- Morais, H.; Caramori, P.H.; Koguishi, M.S.; Ribeiro, A.M.A. (2008) Escala fenológica detalhada da fase reprodutiva de *Coffea arabica*. Bragantia, 67 (1): 257-260.
- Mouradov, A.; Glassick, T.V.; Hamdorf, B.A.; Murphy, L.C.; Marla, S.S.; Yang, Y.; Teasdale, R. (1998) Family of MADS-box genes expressed in early male and female reproductive structures of Monterey Pine. Plant Physiol., 117: 55–61.

- Münster, T.; Deleu, W.; Wingen, L.U.; Ouzunova, M.; Cachárron, J.; Faigl, W.; Werth, S.; Kim, J.T.T.; Saedler, H.; Theißen, G. (2002a) Maize MADS-box genes galore. Maydica, 47:287-301.
- Münster, T.; Faigl, W.; Saedler, H. & Theißen, G. (2002b) Evolutionary aspects of MADS-box genes in the Eusporangiate Fern *Ophioglossum*. Plant biology, 4: 474-483.
- **Ohmori, S.; Kimizu, M.; Sugita, M.; Miyao, A.; Hirochika, H.; Uchida, E.; Nagato, Y.; Yoshida, H.** (2009) *MOSAIC FLORAL ORGANS1*, an AGL6-like MADS box gene regulates floral organ identity and meristem fate in rice. The Plant Cell, 21:3008-3025.
- Oliveira, R.R.; Chalfun-Junior, A.; Paiva, L.V.; Andrade, A.C. (2010) In Silico and Quantitative Analyses of MADS-Box Genes in *Coffea arabica*. Plant Mol. Biol. Rep., 28: 460-472.
- Parenicova, L.; Folter, S.; Kieffer, M.; Horner, D. S.; et al. (2003) Molecular and phylogenetic analyses of the complete MADS-box transcription factor family in Arabidopsis: New openings to the MADS world. The Plant Cell, 15: 1538-1551.
- Pelaz, S.; Ditta, G.S.; Baumann, E.; Wisman, E.; Yanofsky, M.F. (2000) B and C organ identity functions require SEPALLATA MADS-box genes. Nature, 405: 200-203.
- **Pinyopich, A.; Ditta, G.S.; Savidge, B.; Liljegren, S.J.; Baumann, E.; Wisman, E.; Yanofsky, M.F.** (2003) Assessing the redundancy of MADS-box genes during carpel and ovule development. Nature, 424: 85-88.
- Pnueli, L.; Abu-Abeid, M.; Zamir, D.; Nacken, W.; Schwarz-Sommer, Z.; Lifschitz, E. (1991) The MADS box gene family in tomato: temporal expression during floral development, conserved secondary structures and homology with homeotic genes from *Antirrhinum* and *Arabidopsis*. The Plant J., 1: 255– 266.
- Popham, R.A. (1963) Developmental studies of flowering. Columbus, Ohio State University. pp. 138-156.
- **Ptashne, M.** (2004) A genetic switch: phage lambda revisited. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press, 3° ed.
- Ptashne, M. (2005) Regulation of transcription: from lambda to eukaryotes. Trends Biochem. Sci. 30: 275–279.
- Rezaian, M.A. & Krake, L.R. (1987) Nucleic acid extraction and virus detection in grapevine. J. Virol. Meth., 17: 277-285.
- Riechmann, J.L.; Krizek, B.A. & Meyerowitz, E.M. (1996) Dimerization specificity of Arabidopsis MADS domain homeotic proteins *APETALA1*, *APETALA3*, *PISTILLATA* and *AGAMOUS*. Proc. Nat. Acad. Sci., 93: 4793-4798.
- Rijpkema, A.S.; Royaert, S.; Zethof, J.; van der Weerden, G.; Gerats, T. & Vandenbussche, M. (2006) Analysis of the *Petunia TM6* MADS box gene reveals functional divergence within the *DEF/AP3* lineage. The Plant Cell, 18: 1819-1832.
- Rounsley, S.D.; Ditta, G.S. & Yanofsky, M.F. (1995) Diverse roles for MADS box genes in *Arabidopsis* development. The Plant Cell, 7: 1259-1269.
- Rozen, S. & Skaletsky, H.J. (2000) Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds) *Bioinformatics Methods and Protocols: Methods in Molecular Biology*. Humana Press, Totowa, NJ, pp 365-386.
- Schultz, J.; Milpetz, F.; Bork, P. & Ponting, C.P. (1998) SMART, a simple modular architecture research tool: Identification of signaling domains. Proc. Natl. Acad. Sci. USA, 95: 5857–5864.
- Schwarz-Sommer, Z.; Huijser, P.; Nacken, W.; Saedler, H.; Sommer, H. (1990) Genetic control of flower development by homeotic genes in *Antirrhinum majus*. Science, 250: 931-936.

- Sheldon, C.C.; Rouse, D.T.; Finnegan, E.J.; Peacock, W.J.; Dennis, E.S. (2000) The molecular basis of vernalization: the central role of *FLOWERING LOCUS C (FLC)*. Proc. Natl. Acad. Sci. USA, 97: 3753-3758.
- Sitnikova, T.; Rzhetsky, A. & Nei, M. (1995) Interior-branch and bootstrap tests of phylogenetics trees. Mol. Biol. and Evol., 12: 319-33.
- Soltis, D.E., Soltis, P.S., Albert, V.A., Oppenheimer, D.G., de Pamphilis, C.W., Ma, H., Frohlich, M.W. & Theissen, G. (2002) Missing links: the genetic architecture of flower and floral diversification. Trends Plant Sci., 7: 22-31.
- Sommer, H.; Beltrán, J.P.; Huijser, P.; Pape, H.; Lönnig, W.E.; Saedler, H.; Schwarz-Sommer, Z. (1990) Deficiens, a homeotic gene involved in the control of flower morphogenesis in Antirrhinum majus: the protein shows homology to transcription factors. EMBO J., 9: 605-613.
- Sondahl, M.R. & Sharp, W.R. (1979) Research in *Coffea* spp. and applications of tissue culture methods. In: Paddock, E. F.; Raghavan, V. (Eds.). Plant cell and Tissue Culture: Principles and Applications, 527-584.
- Tamura, K.; Dudley, J.; Nei, M. & Kumar, S. (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol. Biol. Evol., 24: 1596-1599.
- **The InterPro Consortium** (1999) The InterPro database, an integrated documentation resource for protein families, domains and functional sites. Nucleic Acids Research, 29 (1): 37-40.
- Theissen, G.; Becker, A.; Di Rosa, A.; Kanno, A.; Kim, J.T.; Munster, T.; *et al.* (2000) A short history of MADS-box genes in plants. Plant Mol. Biol., 42: 115-149.
- **Theissen, G. & Melzer, R.** (2007). Molecular mechanisms underlying origin and diversification of the angiosperm flower. Annals of botany, 100 (3): 603-619.
- Theißen, G. & Saedler, H. (2001) Floral quartets. Nature, 409: 469-471.
- **Thompson, J.D.; Higgins, D.G. & Gibson, T.J.** (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Research, 22: 4673-4680.
- Tröbner, W.; Ramirez, L.; Motte, P.; Hue, I.; Huijser, P.; Lönnig, W.E.; Saedler, H.; Sommer, H.; Schwarz-Sommer, Z. (1992) GLOBOSA: a homeotic gene which interacts with DEFICIENS in the control of Antirrhinum floral organogenesis. EMBO J., 11 (13): 4693-4704.
- Vandenbussche, M.; Zethof, J.; Royaert, S.; Weterings, K. & Gerats, T. (2004). The duplicated B-class heterodimer model: whorl-specific effects and complex genetic interactions in *Petunia* hybrida flower development. The Plant Cell, 16: 741-754.
- Viaene, T.; Vekemans1, D.; Becker, A.; Melzer, S. & Geuten, K. (2010) Expression divergence of the AGL6 MADS domain transcription factor lineage after a core eudicot duplication suggests functional diversification. BMC Plant Biology, 10: 148.
- Vieira, L.G.E.; Andrade, A.C.; Colombo, C.A.; Moraes, A.H.A.; Metha, A.; *et al.* (2006) Brazilian coffee genome project: an EST-based genomic resource. Braz. J. Plant Physiol., 18 (1): 95-108.
- Weigel, D. & Meyerowitz, E.M. (1994) The ABCs of floral homeotic genes. Cell, 78: 203-209.
- Winter, K-U.; Becker, A.; Münster, T.; Kim, J.T.; Saedler, H.; Theißen, G. (1999) MADS-box genes reveal that gnetophytes are more closely related to conifers than to flowering plants. Proc. Nat. Acad. Sci. USA, 96: 7342-7347.
- Winter, K-U.; Saedler, H.; Theißen, G. (2002) On the origin of class B floral homeotic genes: functional substitution and dominant inhibition in *Arabidopsis* by expression of an ortholog from the gymnosperm *Gnetum*. The Plant J., 31: 457-475.
- Wormer, T.M. & Gituanja, J. (1970a) Floral initiation and flowering of *Coffea arabica* L. In Kenya. Expl. Agric., 6: 157-170.
- Wormer, T.M. & Gituanja, J. (1970b) Seasonal patterns of growth and development of *arabica* coffe in Kenya. Kenya Coffee, 35: 270-277.
- Yanofsky, M.F.; Ma, H.; Bowman, J.L.; Drews, G.N.; Feldmann, K.A. & Meyerowitz, E.M. (1990) The protein encoded by the *Arabidopsis* homeotic gene *AGAMOUS* resembles transcription factors. Nature, 346: 35-39.
- Yoo, M.J.; Soltis, P.S.; Soltis, D.E. (2010) Expression of floral MADS-box genes in two divergent water lilies: *Nymphaeales* and *Nelumbo*. Int. J. Plant Sci., 171: 121-146.
- Yu, H.; Xu, Y.; Tan, E.L.; Kumar, P.P. (2002) AGAMOUS-LIKE 24, dosage-dependent mediator of the flowering signals. Proc. Natl. Acad. Sci. USA, 99: 16336-16341.
- Zachgo, S.; de Andrade Silva, E.; Motte, P.; Tröbner, W.; Saedler, H.; Schwarz-Sommer, Z. (1995) Functional analysis of the *Antirrhinum* floral homeotic Deficiens gene in vivo and in vitro by using a temperature-sensitive mutant. Development, 121: 2861-2875.
- Zahn, L.M.; Leebens-Mack, J.; dePamphilis, C.W.; Ma, H. & Theissen, G. (2005) To B or not to B a flower: the role of *DEFICIENS* and *GLOBOSA* orthologs in the evolution of the angiosperms. J. Hered., 96: 225-240.
- Zhang, H. & Forde, B.G. (1998) An Arabidopsis MADS box gene that controls nutrient-induced changes in root architecture. Science, 279: 407-409.

APÊNDICE

Item 1 – Lista de primers

	Direcão	Primers
GAPDH	Forward	GAATTGGACGTTTGGTTGCT
	Reverse	AACTGGCATTGGACACAACA
1 - CaAGL17-1	Forward	CGAGGTGTCCGTATGAGAAAG
	Reverse	GCACCACTCCTTACGAATTGA
2 - CaSVP-1	Forward	TGGATACCTTTGACTCTTGG
	Reverse	CTCTTCTCATTTGCCTCAGTTG
3 - CaTM3-1	Forward	CGTCTCACCTACCCAGCATTTT
	Reverse	GTGCTTGTTCCATTCTTCTGCT
3.1 - CaTM3-2	Forward	GCAGCAAGTAGACAAGTGACA
	Reverse	CTCCCAAAACAATAGGCTCT
4 - CaMADS	Forward	CGCACCTGATCTGCTCATAA
	Reverse	TGGCAATTGGCATAAACATC
5 - CaAP3	Forward	TCAAAAGGCTGTAGGGGTTG
	Reverse	TGTGAAGATTAGGCTGGGTAGG
6 - CaAGL17-2	Forward	AACCCTCTTCAGCAATCACA
	Reverse	GCCATACAGTTCTTCTCCCATC
7 - CaMd	Forward	GTTCTCCCCCTCTGGAAGAC
	Reverse	CCCATTTACTTGTGCTGCTTG
8 - CaSEP4	Forward	TGTGATGCTGAGGTTGCTCT
	Reverse	GGCCTTGTTTACTTCCATCA
9 - CaSEP3	Forward	GCGAAGAAATGGGCTATTGA
	Reverse	ACTTGAGTCCGTGTTGATCTG
10 - CaPI	Forward	TGGAAAGAAGCTGTGGGATG
	Reverse	TGAAGATTTGGCTGCATTGG
11 - CaSEP1/2	Forward	TGGCCCTCATCATCTTCTCT
	Reverse	TTTCCAAGCTCCTTTCTAATGC
12 - CaAG	Forward	AGCAGAGTCCGATCCAAAAA
	Reverse	CTTCGCCTCCACCAACAC
13 - CaAP1	Forward	ATGAGCTCGAAAGACCTCCA
	Reverse	TGGTTTCTCCGCCTTTATC
14 - CaTM6	Forward	ACAATATGTGCCGTCTGCAA
	Reverse	TTCTCATACGAGGACTAGCACA
15 - CaFUL	Forward	GCTGCTGGAGTCTGGAACAT
	Reverse	CATTGGTTTGGCTGCTCTG
16 - CaAGL42	Forward	GCTTGCCTATTTAGAGTCCTTG
	Reverse	GGTCGCTTCTTTGATCTTCCT
17 - CaSVP-2	Forward	TTTTGTGCGATGCTGATGTT
	Reverse	TCTGGAATCACTGGCCTCTT
18 - CaAGL24	Forward	TCCGAGTCGAGAAAAATGGT
	Reverse	ACCCAAGTCTTCCATGCTCA
19 - CaFLC	Forward	CTACGACTTCTGCAGCACCA
	Reverse	AAGGTCCCCAATTCAAGGAT
21 - CaAGL6	Forward	GAGAATGCTGCTGAAAGAGA
	Reverse	GGTGAGAAGTGTGCATAGGA

Item 2 - Lista de sequências MADS-box obtidas no cafeeiro

Legendas:

XXX CDS XXX Domínio MADs-box XXX Domínio K-box Obs.: Região dos primers sublinhada nas sequências

1) CaAGL17-1 (Coffea arabica AGAMOUS-like17-1)

CTCTCTTTTCTCTCTCTCTCTCTCCTCCCCCCCCCTTTGAGAGTTAGCAAGTCCTTTCTCTGTTTATCGCTTCAAAAAGTCCATCTCTA GTGATTAGGAGGATCGATAATTCGACGAGCAGGCAAGTGACGTTTTCCAAGAGGAGGAATGGGTTGCTGAAGAAGGCTAAGGAGTTGGCGATCC TGTGCGACGCGGAGGTTGGAGTCGTCATCTTCTCCAGTACTGGCAAGCTCTATGATTTCTCTAGCTCCAGCATGAAGTCAGTGATTGAACGATA TAACAAAGTGAAAGAGGAGCATCACCAATTGGCCAATCCGAGTTCTGAAGTCAAGTTTTGGCAGGGGAAGTGGCAATGCTCAGGCAACAACTG CAGAACTTGCAAGAGAATCATCGGCAAATGAGGGGCGAAGAGCTTTCTGGTTTGAGTGTCAAAGATCTACAGAACTTGGAGAACCAACTTGAAA TGTTGAACTGTATAAGAAGGTAAACCTAATTCGTGAAGAAAATCAAGAATTGTACAAGAAGGTTTATGGAACCAAGGAGGCGAATGGAACTAAC CGGCAACAGGGGCATCAAAACTGGGATTGCAACTCCATTAGTATGAAGAAAATGCTGAGCTGATTGCTTGTTGATGAAGATCCACCATGCATAC AACACAATGTTCAATTCGTAAGGAGTGGTGCAAAAACCACATATTACTCAGAAAAAGTCTTCAATTTCATTGAAAGGATCTTGACATATCCGAG GTAATTAATGCTATAGCTGTCTATGTAAGATGGTGAGAAATAAAGAAATACTGCTAATTGCTTGTCTCAGACTTTCTTCAGATATCTTTCTATA GGGCATCTCTACTCAATTGAAAAATGTACACAAGCATCTCTACACAATAACTGGGATCATAAAAAATGTACTTGTATAGGTATCCACTTGAGTA AGATGCACCCTGGCGACAAGAAATGCCATGAAAAGTAGAAAGTACAGGATGAAGCAAGTCAAAGAATTTGATGGACCAAACAGTCAATGAAAAA АААААААААААААААААААААААААААААААААААААА

Reads do contig: CA00-XX-BP1-123-A07-AG.F (clone)

>Proteína

MGRGKIVIRRIDNSTSRQVTFSKRRNGLLKKAKELAILCDAEVGVVIFSSTGKLYDFSSSSMKSVIERYNKVKEEHHQLANPSSEVKFWQREVA MLRQQLQNLQENHRQMRGEELSGLSVKDLQNLENQLEMSLRGVRMRKDQVLIDEIQELNRKGSLIHQENVELYKKVNLIREENQELYKKVYGTK EANGTNKTALLTNNLSIREDPDGPVHLQLSQPQQNYETATGASKLGLQLH-

1 81	CTCTCTTTCTCTCTCTCTCTCTTTTCTCCTTCCGCTCTTGAGAGTTAGCAAGTCCTTTCTCTGTTTATCGCTTCAA AAAGTCCATCTCTAGCCCTTTTTTTCCTTCCATCTTTTTTTT	80 160
161	R T S <mark>M G R G K I V I R R I D N S T S R Q V T F S K</mark> TCGGACCAGGATGGGGAGAGGAAGATAGTGATTAGGAGGATCGATAATTCGACGAGGCAAGTGACGTTTTCCAAGA	240
241	<mark>R R N G L L K K A K E L A I L C D A E V G V V I F S S</mark> GGAGGAATGGGTTGCTGAAGAAGGCTAAGGAGTTGGCGATCCTGTGCGACGCGGAGGTTGGAGTCGTCATCTTCTCCAGT	320
321	<mark>T G K L Y D F S S S S M K S V I E R Y</mark> N K V K E E H H ACTGGCAAGCTCTATGATTTCTCTAGCTCCAGCATGAAGTGAAGTGAACGATGAAAGTGAAAGAGGAGCATCA	400
401	Q L A N P <mark>S S E V K F W Q R E V A M L R Q Q L Q N L</mark> CCAATTGGCCAATCCGAGTTCTGAAGTCAAGTTTTGGCAGAGGGAAGTGGCAATGCTCAGGCAACAACTGCAGAACTTGC	480
481	<mark>Q E N H R Q M R G E E L S G L S V K D L Q N L E N Q L</mark> AAGAGAATCATCGGCAAATGAGGGGGGGAAGAGCTTTCTGGTTTGAGTGTCAAAGATCTACAGAACTTGGAGAACCAACTT	560
561	<mark>EMSLRGVRMRKDQVLIDEIQELNRKGS</mark> GAAATGAGCCTT <u>CGAGGTGTCCGTATGAGAAAG</u> GACCAAGTGCTTATTGACGAGATACAAGAACTAAACCGGAAGGGTAG	640
641	<mark>L I H Q E N V E L Y K K V</mark> N L I R E E N Q E L Y K K CCTCATCCATCAAGAAAATGTTGAACTGTATAAGAAGGTAAACCTAATTCGTGAAGAAATCAAGAATTGTACAAGAAGG	720
721	V Y G T K E A N G T N K T A L L T N N L S I R E D P D TTTATGGAACCAAGGAGGCGAATGGAACTAACAAAACTGCACTTCTGACAAACAA	800
801	G P V H L Q L S Q P Q Q N Y E T A T G A S K L G L Q L GGTCCTGTACATCTCCAGCTTAGCCAGCCACAAAACTATGAGACGGCAACAGGGGGCATCAAAACTGGGATTGCAACT	880

	H * Y E E N A E L I A C * * R S T M H T T Q C S I R	
881	CCATTAGTATGAAGAAAATGCTGAGCTGATTGCTTGTTGATGAAGATCCACCATGCATACAACACAATGTTCAATTCGTA	960
961	AGGAGTGGTGCAAAAAACCACATATTACTCAGAAAAAGTCTTCAATTTCATTGAAAGGATCTTGACATATCCGAGGTAATT	1040
1041	AATGCTATAGCTGTCTATGTAAGATGGTGAGAAATAAAGAAATACTGCTAATTGCTTGTCTCAGACTTTCTTCAGATATC	1120
1121	TTTCTATAGGGCATCTCTACTCAATTGAAAAATGTACACAAGCATCTCTACACAATAACTGGGATCATAAAAAAATGTACT	1200
1201	TGTATAGGTATCCACTTGAGTAAGATGCACCCTGGCGACAAGAAATGCCATGAAAAGTAGAAAGTACAGGATGAAGCAAG	1280
1281	TCAAAGAATTTGATGGACCAAACAGTCAATGAAAAAGTGACAATAGTACACGCAATAAGGTAAAAAAAA	1360
1361	ААААААААААААААААААААААААААААААААААААААА	1440
1441	ааааааа 1447	

2) CaSVP-1 (Coffea arabica SHORT VEGETATIVE PHASE-1)

Reads:

CA00-XX-CA1-005-C03-AC.F (clone)

>Proteína

MAREKIQIRKIDNATARQVTFSKRRRGLFKKAEELSVLCDADVALILFSSTGKLFEYSSSSMKEILERHNLHSKNLEKMEQPSLELQLVENSNC SRLSKEVAEKSHQLRQMRREELQGLTVDDLQQLERSLEAGLNRVIEKKGEKIMKEINHLQRKGMQLMEENERLKQQVMEVSNGSGQMAADDSEN VLYEEGQSSESVTNVCSSTGGPQDYESSDTSLKLGLPYSSRLEK-

1 81 161 241	$eq:approx_appr$	80 160 240 320
321	<mark>M A R E K I Q I R K I D N A T A R Q V T F S K R R R G</mark> TGGCTAGAGAGAAGATTCAGATCAGGAAAATCGATAACGCCAGGCCAGGTGACTTTCTCAAAGAGAAGAAGAAGAGA	400
401	L F K K A E E L S V L C D A D V A L I L F S S T G K L CTTTTCAAGAAAGCTGAAGAACTTTCAGTTCTCTGCGATGCTGATGTTGCTCTCTCT	480
481	FEYSSSSMKEILERHNLHSKNLEKME TTTTGAGTATTCCAGCTCCGATAGGGAAATACTCGAAAGGCACAACTTACACTCGAAGAACCTTGAAAAAATGGAGC	560
561	Q P S L E L Q L V E N S N C S R L S K E V A E K S H <mark>Q</mark> AGCCGTCGCTTGAGCTGCAGCTGGTGGAGAATAGCAATTGCTCTCGATTAAGCAAGGAAGTTGCTGAGAAAAGCCAT <u>CAA</u>	640
641	L R Q M R R E E L Q G L T V D D L Q Q L E R S L E A G CTGAGGCAAATGAGAAGAGAAGAGACTCCAAGGATTGACTGTTGATGATCTGCAGCAATTAGAGAGGTCGCTTGAAGCTGG	720
721	<mark>L N R V I E K K G E K I M K E I N H L Q R K G M Q L</mark> ATTGAACCGTGTGATAGAGAAAAAGGGCGAAAAGATAATGAAGGAAATTAATCACCTTCAGCGAAAGGGAATGCAATTAA	800
801	MEENERLKQQVMEVSNGSGQMAADDSE TGGAAGAGAATGAGCGTTTGAAGCAACAGGTGATGGAGGTATCCAATGGGAGCGGGCAAATGGCAGCAGATGATTCAGAG	880
881	N V L Y E E G Q S S E S V T N V C S S T G G P Q D Y E AACGTTCTGTACGAGGAAGGCCAATCATCTGAATCGGTTACTAACGTCTGTAGCTCTACCGGCGGTCCACAAGACTATGA	960
961	S S D T S L K L G L P Y S S R L E K * S S M * H Q N AAGTTCTGATACCTCTCTAAATTGGGGCTGCCATACTCAAGTCGATTGGAAAAGTAAAGCAGTATGTAACACCAAAACG	1040

1041	GGAAATGTGTTGTTGGTGGTCACATTGCATGAAATAAGAAGTAAAAGCTAAC	CAGCAACTTCAGATTCTTGATTCAAAAC	1120
1121	CAAGCATATTTAAAGAGAGGGGAGGGAGTTGATGAGACTGTAGATGATGAAA	CCATAACTATCATTGACAGTAAAAATGG	1200
1201	GATACATCGCTAACCAGTATTTCGAGACTTTCTCATTCAT	ATAACTTTAATAGACATTGCATTTAATA	1280
1281	AATTGAGGCCTCTGCCATGGCTATTCTGTGCTAAAAAAAA	1330	

3) CaTM3-1 (Coffea arabica TM3-1) - sem clone

 $\label{eq:strate} a transform of the strategy of the strateg$

Reads:

CA00-XX-CA1-039-D10-AB.R CA00-XX-LV8-017-E03-JM.F CA00-XX-LV8-020-F03-AC.F CA00-XX-LV8-020-F03-AC.F

>Proteína

MARGKTQLKRIENATSRQVTFSKRRSGLLKKAFELSVLCDAEVALIVFSPKGKLYEFSSSSATSTIQRYQKNIKNLCPSRRMEQAQHFEEEVAI LRKKIEILEETRRRFLGDGLDSSSGDELQQIENRLEKSLSIIRSRKSLLFRERMDHLKEEEKILRKENAELRGKYEEATNWNLSISPQPLPLRQ VKEVETQLFIGPPKS-

Mapa do cDNA:

1	ATAAATATTCCCCTTTTATTTCATTTTTGGAGGGGGGGTCTCTTTCGTGCTTCTCGCGAGTTTACCAAAAGAAGAAAAGGG	80
81	ATGAGGTATTGGGGCCTTGCTAGTCTTTCCTTAATCATTTGTGTTTTGCATATATTCGGTGGTGGGAGATGAGACTCAT	160
161	AAAGACCAAGGAGTTTGAGGAATATCGTTTCTCTGCACATCTTGCCTGAAAAAATCTCTCTTTAGGGTTTCTGCTAG <u>CGT</u>	240
241	<u>CTCACCTACCCAGCATTTT</u> CCAGCTTTTCTCATCTCCCATCTTGCAGATCTGAGACATCCCAGAAAAGGGTTGTAGGTGG	320
	D S I L A L F L L S K F V R <mark>M A R G K T Q L K R I E N</mark>	
321	ATTCAATTTTGGCCCTCTTTTTGTTATCTAAATTTGTAAGAATGGCGAGAGGGAAAACTCAGCTGAAGAGGATTGAAAAT	400
	A T S R Q V T F S K R R S G L L K K A F E L S V L C D	
401	GCAACGAGTAGACAAGTGACATTCTCAAAGAGGAGAAGTGGACTTCTCAAGAAGGCCTTTGAGCTATCAGTTCTTTGTGA	480
4.0.1	A E V A L I V F S P K G K L Y E F S S S S A T S T I	FCO
481	IGCAGAAGIIGCACIGAICGIIIICICCCCAAAAGGGAAGCIIIAIGAGIICICAAGIICCAGIGCCACAAGIACAAIAC	560
EC1	<mark>Q R Y Q K</mark> N I K N L C P S R R M <mark>E Q A Q H F E E V A</mark>	C 1 0
100	AGCGCIACCAGAAGAAIAICAAGAAICIAIGICCG <u>AGCAGAAGAAIGGAACAAGCAC</u> AGCAIIICGAAGAAGAGGGIIGCA	040
C 4 1	I L R K K I E I L E E T R R R F L G D G L D S S S G D	700
041	AICCIGAGAAAAAAAIIGAGAICCIIGAGGAAACICGACGAAGGIICIIGGGAGAIGGIIIAGAIICCICCAGIGGIGA	120
701	ELQQIENRLEKSLSIIRSRKSLLFRE	000
/21	IGAAIIGCAACAGAIAGAAAAICGGCIGGAGAAAAGIIIAAGCAICAICAGGICAAGAAAGA	800
0.01	RMDHLKEEEKILRKENAELRGKYEEAT	000
801	GAATGGATCACCIGAAAGAAGAGAGAGAAAATTCTCAGAAAGGAAAATGCAGAATTGCGGGGAAAGTATGAAGAAGCCACC	880
0.01	NWNLSISPQPLPLRQVKEVETQLFIGP	0.00
σστ	AATTGGAACTTATCGATTAGTCCACAACCCTTGCCACTGAGACAGGTAAAGGAGGTAGAAACACACAC	960
0.61	PKS*NCSPPKKKEQFISLCNFD	
90I	icccaaaagiiagaalig ftcccccccaaagaaaaaagaacaatttatcagcctatgtaacTTTGATT 1028	

3.1) *CaTM3-2* (*Coffea arabica TM3-2*)

Reads:

CA00-XX-CA1-035-H06-QH.F (clone) CA00-XX-RX1-096-D04-EB.F

>Proteína

MVRGKTQIKRIENAASRQVTFSKRRRGLLKKAFELSVLCDAEVALIIFSPSGKLYEFSSSSATSTIERYQKNIRNLCPSEKMALQHSQNFEEEV AILRKKLEILEETKRKLLGDGLDTSSFDELQQIEGQLERSLNIIRSRKSLLFWEQIDHLKEEEKILRKENAELREKMNLQYEQQRLGPSISRQP LSLRQVKEIETRLFIGLPESSNYP-

Mapa do cDNA:

1 81	CTTGCCTGAAAAATCTCTCTTTTAGGGTTTCTACTAGTGTCTTTGCTACCCATTATTTTTCCAGCTTTTCCAACCTACCATC TTGCAGATCTGAGACTTCCCAGAAAAGGGTTGTCGGTGGATTTAATTTTGGACCTCGTCCTTATCTAAAATTGTAAGAA	80 160
161	<mark>M V R G K T Q I K R I E N A A S R Q V T F S K R R R G</mark> TGGTGAGGGGGAAAACTCAGATTAAGAGGATAGAAAAT <u>GCAGCAAGTAGACAAGTGACA</u> TTCTCAAAGAGGAGAAGGGGA	240
241	LLKKAFELSVLCDAEVALIIFSPSGKL CTTCTCAAGAAGGCCTTTGAACTATCAGTTCTTTGTGATGCAGAAGTTGCACTGATCATTTTCTCCCCCAGCGGGAAGCT	320
321	YEFSSSSATSTIERYQKN <mark>IRNLCPSE</mark> TTATGAGTTCTCAAGTTCCAGTGCCACAAGTACAATAGAGCGCTACCAGAAGAATATCAGGAATCTATGTCCGAGCGAAA	400
401	K M A L <mark>Q H S Q N F E E E V A I L R K K L E I L E E T</mark> AAATGGCTCTACAACATTCACAGAATTTTGAAGAAGAGGTTGCAATTCTGAGAAAAAAACTTGAGATTCTTGAAGAAACT	480
481	<mark>K R K L L G D G L D T S S F D E L Q Q I E G Q L E R S</mark> AAACGAAAGTTGTTGGGAGATGGCTTGGATACCTCCTCTTTGATGAGCTGCAACAGATAGAAGGTCAATTGGAGAGAAG	560
561	<mark>LNIIRSRKSLLFWEQIDHLKEEEKIL</mark> TTTAAACATCATTAGGTCAAGAA <u>AGAGCCTATTGTTTGGGAG</u> CAAATTGATCACTTGAAAGAAGAGGGGAGAAAATCCTCA	640
641	<mark>R K E N</mark> A E L R E K M N L Q Y E Q Q R L G P S I S R Q GGAAGGAAAATGCGGAATTACGAGAAAAGATGAATTTGCAGTATGAGCAGCAACGATTGGGACCATCGATTAGTCGACAG	720
	PLSLRQVKEIETRLFIGLPESSNYP*K	
721	CCCTTATCACTGAGACAAGTAAAGGAGATAGAAACACGACTGTTTATTGGACTTCCTGAAAGTTCAAACTATCCCTAGAA	800
801	AAAAAGAAAAAAAAAAAAGAAATTGAACAATTTACCAGTGGAGGTGATTTTCGCATGATTTGTTTTTCTTTACTTCTTGT	880
881 961	cagtitatcititattiticititicitaaaaggcaigtiatGaggcaaaAggcacagcitciatattGatggatgaaatat Tgaataaggctacttgtattgcaaaaaaaaaaaaaaa	960

4) CaMADS (Coffea arabica MADS) - Não finalizada

Reads: CA00-XX-CA1-050-D02-AC.F (clone)

Proteína: MLRGKTQIKRIESATSRQVTLSKRRRGVLKKAFELSQLCDAEVALIGGSFMNSQVPGTFTINIRT

Mapa do cDNA:

- 81 CTAAAAAATCTCTCTTTAGGGTTTCTACTAGTGTCTTGCCTACCCAACATTTTCCAGCTTCTGTCATCTCCCATCTTGCA

161	GATCTGAGACTTCCCAGAAAAGATATTGTCTGGTTTTCCATTTTTTGGGGTCAATCAGTGCCACTCCCGGGGTCGTC	240
	G G F N F A P R F L S T I V R <mark>M L R G K T Q I K R I E</mark>	
241	GGTGGATTTAATTTTGCACCTCGTTTCTTATCTACAATTGTAAGAATGTTGAGAGGGAAAACTCAGATTAAGAGGATAGA	320
	SATSROVTI, SKRRRGVI, KKAFEI, SOI.	
321	AAGTGCAACAAGCAGACAAGTGACATTGTCAAAGAGGAGAAGGGGGAGTTCTCAAGAAGGCCTTTGAGCTATCTCAGCTC	400
	<mark>C D A E V A L I</mark> G G S F M N S Q V P G T F T I N I R	
401	TGTGATGCAGAAGTTGCACTGATCGGGGGTAGCTTCATGAACTCTCAAGTTCCAGGTACATTTACCATCAACATTCGCA	480
481	CCTGATCTGCTCATAATCTCTATCTGTCTCTCTCCAGCTAAACGGTATCATATTAATCAGAATTTTTGTCCACTTTTACA	560
561	ACGGATTGGCCATTAATCATTCCAAGCATGTTCTCAAGCTTTTGTCAGGNCTAAGGGATCAGGNATCCCTTATCTGGGTT	640
641	TATGTCCGTTCTTATGAATCAATTTTACCAATAAGTGCATTTTTTATGAGTTTGACTGAAGGGTTAGATATTA5 714	

5) CaAP3 (Coffea arabica APETALA3)

XXX Splicing alternativo:

poli A ou

Reads:

CA00-XX-AR1-026-G02-EB.F CA00-XX-CB1-011-C02-EQ.F (clone) CA00-XX-CS1-042-G07-JM.F CA00-XX-CS1-076-B12-AB.F CA00-XX-FB2-009-F06-EZ.F CA00-XX-FB2-013-F09-UT.F CA00-XX-FB2-035-H12-BM.F CA00-XX-FB2-092-G05-CE.F

>Proteína

MARGKIQIKRIENQTNRQVTYSKRRNGLFKKAHELTVLCDARVSIIMVSSTQKLHEYISPTATTKQLVDQYQKAVGVDLWSSHHEKMQEQLKKL KEVNRNLRKEIRQRMGESLNDLSYDELGFLIEDVDNSLRAIRERKYKVIGNQIETHKKKVRNVEEIHRNLLLELDARGEDPHYGLVDNGGGDYN PVLGYPRVLALRFQPTQPNLHSGGGSSDLTTFALLE-

1	GCAACGACTGGGACCCAATCACAGCAAGAAGAGTCTTAAAATCTTGAAGCATTGGCAGTGACAGATAACTAAGAAGTGAC	80
81	TGATTCTTAAACAACTTTATCAACTTCTGCCAACACTTTTTCGCGCAATTGGTAAGTCTTTCCATTTTTAGTAACCCATG	160
161	GATCTAAACAATTCTCAACCTCTCCTTTCATTTCAGAAAATCCATGTGCGCTTATCCTCTTCTTGAGTAGGAACAGTAAT	240
241	AATACTGGTAGATGGCTAGATTCATATAACTCAACTGACTATTTAAAACTCTTTAAAGAGGGGCTTAGGCGTAAAGAAGAG	320
	EKKIVV*KGEEEEQIEVVIVLERGQK <mark>M</mark>	
321	AAAAGAAAATAGTCGTATAGAAGGGGGGGGGGGGGGGGG	400
	A R G K I Q I K R I E N Q T N R Q V T Y S K R R N G L	
401	GCTCGTGGGAAGATCCAGATCAAGCGGATAGAGAATCAAACCAACAGGCAAGTGACCTACTCTAAGAGAAGAAACGGGCT	480
	F K K A H E L T V L C D A R V S I I M V S S T Q K L	
481	GTTTAAGAAAGCTCATGAGCTCACCGTCCTGTGCGATGCTAGAGTTTCCATCATGGTCTCCAGTACCCAGAAGCTTC	560
	<mark>H E Y I S P</mark> T A T T K Q L V D Q <mark>Y Q K A V G V D L W S</mark>	
561	ACGAATACATCAGTCCCACAGCCACGACAAAGCAGTTGGTTG	640

	S H H E K M Q E Q L K K L K E V N R N L R K E I R Q R	
641	TCCCACCACGAGAAAATGCAAGAGCAATTGAAGAAGCTAAAGGAGGTGAATAGAAATCTTCGTAAGGAGATCAGGCAGAG	720
721	M G E S L N D L S Y D E L G F L I E D V D N S L R A GATGGGGGAGAGTTTGAATGATCTGAGTTATGATGAGGCTCGGTTTTCTCATAGAAGACGTGGATAATTCTCTGAGGGCCA	800
801	<mark>I R E R K Y K V I G N Q I E T H K K K V R N V E E I H</mark> TTCGGGAGAGGAAGTACAAGGTAATTGGGAACCAGATCGAGACTCACAAGAAAAAGGTGAGGAATGTTGAAGAAATACAC	880
881	<mark>RNLLELDA</mark> RGEDPHYGLVDNGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	960
961	PVLGYPRVLALRFQPTQPNLHSGGGSGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	1040
1041 1121	S D L T T F A L L E * Q F D V L Y I Y I * S L F L R C CGGATCTTACGACTTTTGCCCTGCTTGAGTAGCAATTTGATGTACTATATATA	1120

6) CaAGL17-2 (Coffea arabica AGL17-2)

Reads: CA00-XX-CB1-059-H08-BG.F CA00-XX-CS1-036-A09-UT.F (clone) CA00-XX-RT8-023-H01-BM.F

>Proteína

MGRGKIVIRRIDNSTSRQVTFSKRRNGLLKKAKELAVLCDAEVGVVIFSSTSKLYEYANTSMKSVLERYSKAKEERHQLLSPLSEVKFWQREAT ILRQQLHNLQEIHRQLMGEELYGLSVKDLQGLENQLEMSLRGIRMKKEQILTDEIRELHRKGCLIHQENAELYKKVNFHQQEIISLHKKAYSTS NSNATHGNTITPYGFAITEEQHAPIHLQLSQPESQNFVTSEGTSESR-

R R I D N S T S R Q V T F S K R R N G L L K K A K E 160 81 CCGGAGGATCGATAACTCGACAAGCAAGCAAGGAGAGATTCTCTCTAAGAGAAGAAATGGGCTGTTGAAGAAGGCCAAGGAGC 160 161 L A V L C D A E V G V V I F S S T S K L Y E Y A N T S TGGCTGTTCTCTGTGATGCAGAAGTTGGAGTTGTCATCTTCTCAAGCAAG	1	I T L F S N H S Y K L I K D R A G N I <mark>M G R G K I V I</mark> AT <u>AACCCTCTTCAGCAATCACA</u> GCTATAAGCTCATCAAAGATCGAGCAGGGAATATTATGGGAAGGGGGAAGATTGTGAT	80
LAVLCDAEVGVVIFSSTSKLYEYANTS 161 TGGCTGTTCTCTGTGATGCAGAAGTTGGAGTTGTACTCTCTCAAGCACTAGCAAGCTCTATGAATATGCAAACACCAGC 240 MKSVLERYSKAAGTGGAAGTTGGAGTTGTCATCTTCTCAAGCACTAGCAAGCTCTATGAATATGCAAACACCAGC 240 MKSVLERYSKAAGCGAAAGCGAAAGCGAAAGCGAAAGCAGAAGCGCCATCAACTCCTTAGCACACTATCGGAGGTCAAGGT 320 MQREATILERQQQLHNLQEIHRQLACACTGGAGGAAATTGACAGGGAGAAGAAC 400 11 TGGCAAAGGGAGGCAACAATCCTCCGGCAACAATTACACAACTTGCAGGAAATTCACAGGCAAATTGGAGGAAGAAC 400 121 TGGCCAAAGGGAGGCAACAATCCTCCGGCAACAATTACACAACTTGCAGGAAAATTCACAGGGCAACTGGAGGAAGAAC 400 401 LYGLSVKDLQGLENQQCAACAATTACACAAGGTCTAGAGAAACACCAACTTGAAAATGAGGTTTAAGGGGCATCCGTATGAAAAAAG 480 401 TGTATGGCCTGAGCGTTAAAGAACCTACAAGGTCTAGAGAAACCAACTTGAAAATGAGGTTTAAGGGGCATCCGTATGAAAAAAGGCGGAACTTTATAA 560 640 GAACAAATACTGGATGAGAATTCGGGAAATTACTCATGGATAAGGGCTGCCTCATTCAT	81	<mark>R R I D N S T S R Q V T F S K R R N G L L K K A K E</mark> CCGGAGGATCGATAACTCGACAAGCAGGCAAGTGACTTTCTCTAAGAAAATGGGCTGTTGAAGAAGGCCAAGGAGC	160
M K S V L E R Y S K A K E E R H Q L L S P L S E V K F 241 ATGAAATCCGTCCTGGAACGTTACAGCAAAGCGAAAGAAGAGGCGCCATCAACTCCTTAGTCCACTATCGGAGGTCAAGTT 320 321 M Q R E A T I L R Q Q L H N L Q E I H R Q L M G E E 400 321 TTGGCAAAGGGAAGCAATCCTCCGGCAACAATTACACAACTTGCAGGAAATTCACAGGCAATT <u>GATGGGAGAAAAAC</u> 400 401 L Y G L S V K D L Q G L E N Q L E M S L R G I R M K K 480 481 GAACAAATACTGACGATGAGAATTCGGGAACTACATCGTAAGGGCTGCCTCATTCAT	161	<mark>L A V L C D A E V G V V I F S S T S K L Y E Y A N T S</mark> TGGCTGTTCTCTGTGATGCAGAAGTTGGAGTTGTCATCTTCTCAAGCACTAGCAAGCTCTATGAATATGCAAACACCAGC	240
W Q R E A T I L R Q Q L H N L Q E I H R Q L M G E E 321 TTGGCAAAGGGAGGCAACAATCCTCCGGCAACAATTACACAACTTGCAGGAAATTCACAGGCAATT <u>GATGGGAGAAGAAC</u> 400 401 L Y G L S V K D L Q G L E N Q L E M S L R G I R M K K 480 401 TGTATGGCCTGAGCGTTAAAGACCTACAAGGTCTAGAGAACCAACTTGAAATGAGTTTAAGGGGCATCCGTATGAAAAAG 480 481 GAACAAATACTGACTGATGAGATTCGGGAACTACATCGTAAGGGCTGCCTCATTCAT	241	MKSVLERYSKAKEERHQLLSPL <mark>SEVKF</mark> ATGAAATCCGTCCTGGAACGTTACAGCAAAGCGAAAGAAGAGCGCCATCAACTCCTTAGTCCACTATCGGAGGTCAAGTT	320
L Y G L S V K D L Q G L E N Q L E M S L R G I R M K K 401 <u>TGTATGGC</u> CTGAGCGTTAAAGACCTACAAGGTCTAGAGAACCAACTTGAAATGAGTTTAAGGGGCATCCGTATGAAAAAG 480 481 E Q I L T D E I R E L H R K G C L I H Q E N A E L Y K GAACAAATACTGACTGATGAGATTCGGGAACTACATCGTAAGGGCTGCCTCATTCAT	321	<mark>W Q R E A T I L R Q Q L H N L Q E I H R Q L M G E E</mark> TTGGCAAAGGGAGGCAACAATCCCCGGCAACAATTACACAACTTGCAGGAAATTCACAGGCAATT <u>GATGGGAGAAGAAC</u>	400
EQILTDEIRELHRKGCLIHQEN 481 GAACAAATACTGACTGATGAGATTCGGGAACTACATCGTAAGGGCTGCCTCATTCAT	401	<mark>LYGLSVKDLQGLENQLEMSLRGIRMKK</mark> <u>TGTATGGC</u> CTGAGCGTTAAAGACCTACAAGGTCTAGAGAGGTTTAAGGGGGCATCCGTATGAAAAAG	480
K V N F H Q Q E I I S L H K K A Y S T S N S N A T H 561 GAAGGTAAACTTCCATCAACAAGAAATTATCTCATTGCATAAG G N T I T P Y G F A I T E E Q H A P I H L Q L S Q P E 641 GGAATACCATCAACTATGGATTGCAATTACTGAGGAGCAACATGCCCAATCCCAATTAGCCAGCC	481	<mark>E Q I L T D E I R E L H R K G C L I H Q E N</mark> A E L Y K GAACAAATACTGACTGATGAGATTCGGGAACTACATCGTAAGGGCTGCCTCATTCAT	560
G N T I T P Y G F A I T E E Q H A P I H L Q L S Q P E 641 GGAATACCATAACTCCATATGGATTTGCAATTACTGAGGAGCAACATGCTCCAATCCAACTTAGCCAGCC	561	K V N F H Q Q E I I S L H K K A Y S T S N S N A T H <mark>GAAGGTAAACTTCCATCAACAAGAAATTATCTCATTGCATAAG</mark> AAGGCTTATAGCACGAGTAATTCCAATGCCACACATG	640
	641	G N T I T P Y G F A I T E E Q H A P I H L Q L S Q P E GGAATACCATAACTCCATATGGATTTGCAATTACTGAGGAGCAACATGCTCCAATCCATCTCCAACTTAGCCAGCC	720

	S	Q	Ν	F	V	Т	S	Ε	G	Т	S	Ε	S	R	*	L	Ν	Κ	G	S	Κ	F	S	S	Y	Η	Е	
721	TCA	CAG	AAC	TTC	GTGA	ACGI	ГСА	GAG	GGA	ACC	ГСА	GAA	TCA	AGG	TAA	TTA	AAAC	CAAA	GGC	AGT	AAA	TTC	TCA	TCG	TAT	CAC	GA	800
801	GTC	AAG	ACC	CAA	AGGI	AGG	CAT	GAC	CGG	TTA	GAC	TAC.	ATC	AAC	CAA	CAP	AGAA	AAA	TGC	TAA	CAA	ATG	ATT	TAC.	AGA	GCT	ΤT	880
881	CTT	CAA	TTC	TCT	CGG	GTTC	GΤΑ	ATG	[GT]	TTG	CAT	TTC	GCA	TGG	ГGС	TΤΖ	AACI	GGT	TAT	ATA	AAC	ТСТ	TTA	TAA	GGT	TAT	CC	960
961	ACA	AAT	ACT	AGT	AAC	CAAC	CGC	TGC	CTA	CTG	TAT	TCA	AGT	GCA	AGA	TTC	GCAI	ATG	CCT	CGA	CTG	AAC	TTA	ATG	TGA	ATA	TΤ	1040
1041	TTG	GTC	GTG	AAA	GCT	GAAA	AAA	AAA	AAA	AAA	AAA	A	10	74														

XXX Splicing alternativo: existem as duas proteínas, com e sem essa região.

7) CaMd (Coffea arabica Md)

AAGATCGGTACTGCCTGCAGGTACCGGTCCGGGAATTCCCCGGGCTCACCGCTAACTTTATCTCTCACCCTCTCGTCATCTTCTGAACAGCGGCCT GTTTCGCATATGCAATTAATGACTTTTTTTTTTTTTATCTATGTATAAACAAAAGTACTAGGATACTAGTATAATTTCAGGTTCCATTTAAGAGCTGGC TATTTTAGAGGAGAGAAAACCTAGGAGGAGGAGAAAAATATGGGAAGAGTGAAGCTCCAGATCAAGAAAATCGAGGAGCACAAATAGGCAGGT GACTTTCTCAAAAAGAAGAAATGGGCTTATCAAGAAAGCTTATGAACTTTCTGTGCTCTGTGATGTTGATGTAGCTCTCATCATGTTCTCCCCC TCTGGAAGACTCAGCGTATTTTCAGGAAACAAAAGCCTTGAGGAAATTATGGCACGATATCTGAATCTTCCTGAGCATGAGCGAGGACGGCTGC ATAACCAAGAGTATCTGGAAAAAGCACTTGGCAAGTTGAAATCTGAAGCAGAACTAATCAAGATGTCAGCCCAGTAAGTGTTGACTCTCA ATCAATACGATTTGTGAGGCTGAATACAGGGAACAAATTCTTGAGGAGACTCTGAATCAGGTTCGAGCACGCAAGGTAAATTTGGCTCAACAAG CAAAGTCATATGGATACAGCAGCAGACATGCAGCAGCAGCTGAATGATCATAACAGTCCAATCAGCAAGATCGAGAATGACCCCACAAGTACTGC AACGTCCCCATGATGATCACTTTGGACAAATCATCGACGTAAATCTCTCGCCATGGACGCACCTATATCCAACAGGAACTGATCCGTTTCCAGC TGGACAACCCAGGGAAGGAGCACTCTTGGAGCTGTTCTTATCTCAGCTTGCTCCTGTAAATCAAGATCAGCTTGATAATCTCCCAGTGATTTGTT TGTATTTAGTTTTTTTAATACTATTTTTATTTGATAAGCCTTGTAAGGAGTTAGAATTGTAGAATATTTGAGGTTTTCATTAATAGTTATGT

Reads: CA00-XX-CA1-029-F11-EP.F CA00-XX-CS1-055-D08-EQ.F (clone) CA00-XX-SI3-056-C03-EM.F CA00-XX-SI3-080-C09-EM.F

>Proteína

MGRVKLQIKKIESTTNRQVTFSKRRNGLIKKAYELSVLCDVDVALIMFSPSGRLSVFSGNKSLEEIMARYLNLPEHERGRLHNQEYLEKALGKL KSEADRTNQDVSPVSVDSQIEEIQQEILRYKSQMEDMEKKLRIYEGDPWEINTICEAEYREQILEETLNQVRARKVNLAQQAAQVNGFTTRSAS SILDWFPHQRDQDQIPMLNFLDPTGLIPLRAGQADQRIENMVPASLTLPPQSHMDTAADMQQQLNDHNSPISKIENDPQVLQRPHDDHFGQIID VNLSPWTHLYPTGTDPFPAGQPREGALLELFLSQLAPVNQDQLDNLQ-

1	AAGATCGGTACTGCCTGCAGGTACCGGTCCGGAATTCCCGGGCTCACCGCTAACTTTATCTCTCACCCTCTCGTCATCTT	80
81	CTGAACAGCGGCCTAACGGTGCCAAAAAGCCCCTTCCGATCCGAAATTTGTTACTACCTCTCTTAGCTTCTCTCTC	160
161	CTTCACTTTCCAACTTCTTACATCAACTGTTTCGCATATGCAATTAATGACTTTTTTTT	240
241	CTAGGATACTAGTATAATTTCAGGTTCCATTTAAGAGCTGGCAGGGACAGACTTGATCCTGGCTGAGATTACTACTATAT	320
321	TTGGCTTCCTTCACAAAAAAGGGCTTTCAAGTTTCAACAAGTTCGACAGAGACGTATATTTTAGAGGAGAGGAAACCTAG	400
	G G G K N <mark>M G R V K L Q I K K I E S T T N R Q V T F S</mark>	
401	GAGGAGGGAAAAATATGGGAAGAGTGAAGCTCCAGATCAAGAAAATCGAGAGCACAACAAATAGGCAGGTGACTTTCTCA	480
	K R R N G L I K K A Y E L S V L C D V D V A L I M F S	
481	AAAAGAAGAAATGGGCTTATCAAGAAAGCTTATGAACTTTCTGTGCTCTGTGATGTTGATGTAGCTCTCATCAT <u>GTTCTC</u>	560
	<mark>PSGRLSVFSGNKSLEEIMARY</mark> LNLPE	
561	CCCCTCTGGAAGACTCAGCGTATTTTCAGGAAACAAAAGCCTTGAGGAAATTATGGCACGATATCTGAATCTTCCTGAGC	640
	H E R G R L H N O E Y L E K A L G K L K S E A D R T N	
641	ATGAGCGAGGACGGCTGCATAACCAAGAGTATCTGGAAAAAGCACTTGGCAAGTTGAAATCTGAAGCAGAACAGAACTAAT	720
721	CAAGATGTCAGCCCAGTAAGTGTTGACTCTCAAATTGAGGAAATTCAGCAAGAAATTCTTAGGTATAAGTCCCAAATGGA	800
801	GGATATGGAGAAAAAATTGAGAAATTTATGAAGGTGATCCTTGGGAGATCAATACGATTTGTGAGGCTGAATACAGGGAAC	880
881	AAATTCTTGAGGAGACTCTGAATCAGGTTCGAGCACGCAAGGTAAATTTGGCTCAA <u>CAAGCAAGCAAGTAAATGGG</u> TTT	960
	ттредееттры ериороротры тирт	
	TIK94991TPM TEHŐKDŐDŐT EMTNET	

961	ACTACGAGAAGTGCAAGCAGTATATTGGATTGGTTTCCTCATCAAAGAGACCAAGATCAAATTCCAATGCTCAATTTCTT	1040
1041	D P T G L I P L R A G Q A D Q R I E N M V P A S L T GGATCCCACAGGGCTTATTCCTCTTAGAGCTGGGCAAGCTGATCAGCGCATAGAGAACATGGTACCTGCTTCCTTAACTC	1120
1121	L P P Q S H M D T A A D M Q Q Q L N D H N S P I S K I TTCCTCCTCAAAGTCATATGGATACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG	1200
1201	E N D P Q V L Q R P H D D H F G Q I I D V N L S P W T GAGAATGACCCACAAGTACTGCAACGTCCCCATGATGATCACTTGGACAAATCATCGACGTAAATCTCTCGCCATGGAC	1280
1281	H L Y P T G T D P F P A G Q P R E G A L L E L F L S GCACCTATATCCAACAGGAACTGATCCGTTTCCAGCTGGACAACCCAGGGAAGGAGCACTCTTGGAGCTGTTCTTATCTC	1360
1361 1441 1521	\mathbb{Q} L A P V N Q D Q L D N L Q * F V C I * F F F N T I F AGCTTGCTCCTGTAAATCAAGATCAGCTTGATAATCTCCAGTGATTTGTTTG	1440 1520

8) CaSEP4 (Coffea arabica SEPALLATA4)

 ${\tt cccataaaaatccaatcttggtggcattttgttacatgtttgcttgttttgttaagattttgatagaaccataacactttttttgtgaagaaa$ AAAAGGATTGTAATTCTTTTCTGAGTATTATCTGTTGGTTTTTGATACGGTGTTACAGATACGTTACATAAATTAACACAAAAACAGGATAATGG GAAGGGGTAAAGTGGAGCTGAAGAGAATAGAGAACAAGATAAATAGACAGGTAACCTTTGCAAAAAGAAGGAATGGTCTGCTGAAGAAGGCCTA TGAACTCTCCATCCTATGTGATGCTGAGGTTGCTCTCATCGTATTCTCCAATCGTGGCAAACTCTATGAGTTCTGCAGCAGCTCTAGTATGTCC AAAACGCTGGAGAGGTATCACAGATGCAGCTATGCTGATGCTGGAATGAACCAATCCTCCAAGGATCCACAGGGTGACTACCAGGAGTATCTGA AGCTTAAAGCAAAAGTGGAGGTATTACAACAATCTCAGAGGCATCTTCTGGGGGAAGACTTGGCTCAGCTGGGTGCAAAGCAGCTCGATCAGCT CGAGCGTCAGCTTGATGCATCTTTGCGGCAAATTAGGTCCACAAAGACTCAGCATATGCTGGATCAACTCTCAGATCTTCAGCAAAAGGAAAAA TCACTGATGGAAGTAAACAAGGCCTTGAGGAACAAGCAGTTAGAAGAAACCACTGCAGCATATCAGTTATCATGGGACGTCTCAGAGGAGCATA ATTTACGACATAGATCTCAGACAATTCATCCTGAGGGATTCTTCCAGCCCCTTGAATGCAACAGCTCCATCATGAACTACAACATGGTGGTAGC ACACAATTTGAATTGGAAGGCCAACAAAT<mark>TAAT</mark>ACGTGTGGTACACAGGTTCCTAATTTTGCTGTATTTTCCATTGAAATGGGAAGATAAAAAT AAATTATGTTCTTGGGCCAACAGG<mark>AGG</mark>CATAAGGGGATCCTGTTGGAGTAATGTACGTACCATAGCAGCATTAACATATTGCCAACTCTACAGT ATAATCTTGTGGTGGGGGGGGGGGCTATCTTTATTGAACTGAAATTTGCTTGGACAAATATTATTGTATGACGCAAAACAAAATGTGCAATCAGGC АААААААААААААААААААА

XXX Regiões de splicing alternativo

Reads:

CA00-XX-FB1-109-G09-BG.F (clone) CA00-XX-FB3-000-D06-EZ.F CA00-XX-FR2-014-C01-BF.F CA00-XX-FR2-026-A09-QH.F CA00-XX-RX1-094-G07-EB.F

>Proteína

MGRGKVELKRIENKINRQVTFAKRRNGLLKKAYELSILCDAEVALIVFSNRGKLYEFCSSSSMSKTLERYHRCSYADAGMNQSSKDPQGDYQEY LKLKAKVEVLQQSQRHLLGEDLAQLGAKQLDQLERQLDASLRQIRSTKTQHMLDQLSDLQQKEKSLMEVNKALRNKQLEETTAAYQLSWDVSEE HNLRHRSQTIHPEGFFQPLECNSSIMNYNMVVADAEAEPTQNPSGILPGWML-

Mapa	do	cDNA:
------	----	-------

1	CCAGAAAAGAGTAACTCTTCCTGCCCTAAAAGAGAAAAAGAGACAACCGGTTTGGCAGTGTACTTGTCCATATATAT	80
81	GTTTTCCTTATTTCCCCATAAAAATCCAATCTTGGTGGCATTTTGTTACATGTTTGCTTGTTTTGTTAAGATTTTGATA	160
161	GAACCATAACACTTTTTTTGTGAAGAAAAAAAGGATTGTAATTCTTTTCTGAGTATTATCTGTTGGTTTTTGATACGGTG	240
	V T D T L H K L T Q N R I <mark>M G R G K V E L K R I E N K</mark>	
241	TTACAGATACGTTACATAAATTAACACAAAACAGGATAATGGGAAGGGGTAAAGTGGAGCTGAAGAGAATAGAGAACAAG	320
	I N R Q V T F A K R R N G L L K K A Y E L S I L C D A	
321	ATAAATAGACAGGTAACCTTTGCAAAAAGAAGGAATGGTCTGCTGAAGAAGGCCTATGAACTCTCCATCCTATGTGATGC	400
	E V A L I V F S N R G K L Y E F C S S S M S K T L	
401	TGAGGTTGCTCTCATCGTATTCTCCAATCGTGGCAAACTCTATGAGTTCTGCAGCAGCTCTAGTATGTCCAAAACGCTGG	480
	<mark>ERY</mark> HRCSYADA <mark>GMNQSSKDPQGDYQEY</mark>	
481	AGAGGTATCACAGATGCAGCTATGCTGATGCTGGAATGAACCAATCCTCCAAGGATCCACAGGGTGACTACCAGGAGTAT	560

561	L K L K A K V E V L Q Q S Q R H L L G E D L A Q L G A CTGAAGCTTAAAGCAAAAGTGGAGGTATTACAACAATCTCAGAGGCATCTTCTGGGGGAAGACTTGGCTCAGCTGGGTGC	640
641	K Q L D Q L E R Q L D A S L R Q I R S T K T Q H M L AAAGCAGCTCGATCAGCTCGAGCGTCAGCTTGATGCATCTTGCGGCAAATTAGGTCCACAAAGACTCAGCATATGCTGG	720
721	<mark>D Q L S D L Q Q K E K S L M E V N K A L R N K Q L E</mark> E ATCAACTCTCAGATCTTCAGCAAAAAGGAAAAATCAC <u>TGATGGAAGTAAACAAGGCC</u> TTGAGGAACAAGCAGTTAGAAGAA	800
801	T T A A Y Q L S W D V S E E H N L R H R S Q T I H P E ACCACTGCAGCATATCAGTTATCATGGGACGTCTCAGAGGAGCATAATTTACGACATAGATCTCAGACAATTCATCCTGA	880
881	G F F Q P L E C N S S I M N Y N M V V A D A E A E P GGGATTCTTCCAGCCCCTTGAATGCAACAGCTCCATCATGAACTACAACATGGTGGTAGCCGATGCTGAGGCAAGAGCCAA	960
961 1041 1121 1201 1281 1361 1441	T Q N P S G I L P G W M L * E Y I Y I Y I C S N * T Q CACAAAATCCCAGTGGAATCCTACCAGGATGGATGGATGCTTTGAGAATATATAT	1040 1120 1200 1280 1360 1440 1520
± J Z ±	1991 1927	

9) CaSEP3 (Coffea arabica SEPALLATA3)

```
XXX Região de splicing alternativo
```

Reads:

CA00-XX-FB1-089-H01-EQ.F CA00-XX-FB1-127-G10-AG.F (clone) CA00-XX-FB2-003-E02-CC.F CA00-XX-FB2-030-B02-JF.F CA00-XX-FB2-068-D09-AC.F CA00-XX-FB4-003-A02-JM.F CA00-XX-FR2-091-D09-AC.F

>Proteína

MGRGRVELKRIENKINRQVTFAKRRNGLLKKAYELSVLCDAEVALIIFSNRGKLYEFCSSSRVTLYSMLKTLERYQKCNYGAPEPNISTREALE LSSQQEYLKLKARYEALQRSQRNLLGEDLGPLNSKELESLERQLDMSLKQIRSTRTQVMLDQLTDLQRKEHALNEANKTLKQRLMEGNQVNLQW NPNAQDVGYGRQPAHAQGDGFFHPLDCEPPLQIGYQNDPITVAAAGPSVNNYMAGWLP-

1	GTT	rgag	GAGT	GGG	AAA	TTG	IGTA	ATCI	GTC	TCA	GGG	GCCC	CGCC	CTI	[AA]	ATAA	TTT	TTC	CTC.	ATC	TAA.	ACT	GTO	GGGG	GAT	GTA	TA	80
81	AAC	AGGA	ACCG	AAT	AGT	TAT	CTAC	GTAC	CATG	TGT	AGA	AGA	ATTA	CAI	GTA	ATTO	TGG	GAI	ГGТ	GCA	TAT.	AAA	AAG	GAGO	GA.	AAC.	AA	160
161	AAG	GGAA	GGA	IGCA	GTT	TTGA	AGGC	GGGA	AAC	AAG	ACT	CTG	STTG	TTO	GTG	GTTG	GTAA	TTI	ГGТ	GAG	TTT	CAG	TGG	GTAA	TA.	AGA.	AC	240
	Е	R	Е	V	G	Е	R	М	G	R	G	R	V	E	L	K	R	I	Е	Ν	Κ	Ι	Ν	R	Q	V		
241	GGA	AGA	GAG	GTA	GGA	GAG	AGGI	ATGO	GAA	GAG	GTA	AGGG	STGG	AGC	CTA	AGA	AGGA	TAC	GAG.	AAC	AAG.	ATC	AAI	AGG	GCA.	AGT.	AA	320
	ΤH	r A	K	R	R	Ν	G	L	L	Κ	Κ	А	Y	Е	L	S	V	L	С	D	A	E	J	7 P	1	L	I	
321	CTT	ГТGC	СТАА	GCG	AAG.	AAA	rggo	GCTA	ATTG	AAG	AAA	AGCI	TAT	GAG	GCT	CTCA	GTI	CTA	ATG	TGA	TGC	TGA	.GGI	TGC	TT	TGA	ТС	400
	I	F	S	N	R (G I	K I	L Y	E E	F	' C	: s	s s	5	5 E	R I	νі	l I		Y	S I	М	L	K	Т	L	E	
401	ATC	ГТСТ	CCA	ATA	GAG	GAA	AGCI	ГСТА	ATGA	GTT	TTG	GCAG	STAG	CTC	CTA <mark>(</mark>	GAGI	CAC	GCI	IGT.	ACA	GCA	TGC	TCA	AGA	CG	СТА	GA	480

481	RYQKCNYGAPEPNISTREALEL <mark>SSQQ</mark> GAGGTATCAGAAATGCAACTATGGAGCACCTGAGCACCTAGAGCACCGGGAAGCACTGGAGCTGAGTAGTCAGCAGG	560
561	<mark>EYLKLKARYEALQRSQRNLLGEDLGPL</mark> AGTATTTGAAGCTTAAAGCACGTTACGAAGCTCTACAAAGGAATCTTTTGGGCGAGGACCTTGGTCCTTTA	640
641	NSKELESLERQLDMSLKQIRSTRTQVM AACAGCAAGGAACTTGAATCATTGGAAAGGCAGCTTGATATGTCTCTGAAGCAAAT <u>CAGATCAACACGGACTCAAGT</u> AAT	720
721	<mark>L D Q L T D L Q R K E H A L N E A N K T L K Q R L M</mark> GCTGGATCAGCTCACTGATCTCCAGAGAAAGGAACATGCTTTGAACGAGGCAAACAAA	800
801	$\ensuremath{ {\bf E} }$ G N Q V N L Q W N P N A Q D V G Y G R Q P A H A Q G AAGGAAACCAAGTAAATCTCCAGTGGAATCCGAATGCACAGGATGTTGGGTACGGCCGACAACCAGCTCATGCTCAGGGT	880
881	D G F F H P L D C E P P L Q I G Y Q N D P I T V A A A GATGGTTTTTTCATCCATTGGATTGTGAACCGCCATTACAGATTGGATACCAGAATGATCCAATAACAGTGGCAGCGGC	960
961 1041 1121 1201	G P S V N N Y M A G W L P * S E E K I A C H A N D S AGGCCCGAGTGTGAATAACTACATGGCAGGTTGGTTACCATGATCAGAAGAAAAATTGCCTGTCATGCTAATGATTCCT TTCATAAGACTCGATGCAATTGATGTTCTATGTATTTTTATTTTGACTAAAGAACTTGTGTAATAGCATGAATTTTCATA AGACGCTATTATTATAATCTCCCTAATGTGGACGAACGTTAAGTCATATATTTACTTCGATAAAAAAAA	1040 1120 1200

10) CaPI (Coffea arabica PISTILATA)

Reads: CA00-XX-FB1-008-C09-QH.F CA00-XX-FB1-044-H08-JM.F CA00-XX-FB1-054-E02-JF.F CA00-XX-FB1-073-G10-BM.F CA00-XX-FB1-132-G09-AG.F CA00-XX-FB2-005-C04-AC.F CA00-XX-FB2-006-F10-BF.F (clone) CA00-XX-FB2-009-C06-EZ.F CA00-XX-FB2-035-H08-BM.F CA00-XX-FB2-043-D07-UT.F CA00-XX-FB2-043-H01-UT.F CA00-XX-FB2-058-A07-E0.F CA00-XX-FB2-058-F02-EQ.F CA00-XX-FB2-063-D03-SB.F CA00-XX-FB2-072-B06-BG.F CA00-XX-FB4-002-E05-JE.F CA00-XX-FB4-011-E07-OH.F CA00-XX-FB4-041-G07-AG.F

>Proteína

MGRGKIEIKRIENTNNRHVTYSKRRTGIMKKAKEITVLCDAKVSLIIFGTSGKMHEYISPSTNLVEMLDAYQRSTGKKLWDAKHENLSNEIDRV KKENDSMQIELRHLKGEDITSLNYKELMILEDALENGLAGLREKQSEIIKMIRKTGEMLEDENKQLQYIWHQQEMANMKGAIGERDDVYQRVRD YPSQMPFAFRVQPMQPNLHERI-

Mapa do cDNA:

1

81	<mark>G K I E I K R I E N T N N R H V T Y S K R R T G I M</mark> AGGTAAGATTGAGATCAAGAGGATTGAGAACAACAACAGGCATGTGACTTACTCAAAGAGAAGAACTGGGATCATGA	160
161	<mark>K K A K E I T V L C D A K V S L I I F G T S G K M H E</mark> AGAAGGCTAAGGAGATCACAGTTCTTGGATGCTAAGGTTTCCCTCATCATCTTCGGTACTTCTGGAAAGATGCATGAA	240
241	<mark>Y I S</mark> P S T N L V E M L D A <mark>Y Q R S T G K K L W D A K</mark> TATATAAGTCCTTCAACAAATTTGGTTGAAATGTTGGATGCTTACCAGAGGTCTAC <u>TGGAAAGAAGCTGTGGGATG</u> CTAA	320
321	<mark>H E N L S N E I D R V K K E N D S M Q I E L R H L K</mark> GCATGAGAATTTAAGCAATGAAATTGATAGAAGTTAAGAAAGA	400
401	<mark>G E D I T S L N Y K E L M I L E D A L E N G L A G L R</mark> GGGAAGACATCACATCTTTGAATTACAAAGAACTCATGATTTTAGAAGATGCCCTAGAAAATGGGCTTGCAGGTTTACGC	480
481	<mark>E K Q S E I I K M I R K T G E M L E D E N K Q L Q Y I</mark> GAGAAACAGAGTGAGATCATCAAGATGATCAGGAAAACTGGGGGAAATGCTGGAGGACGAGAATAAGCAGCTTCAATACAT	560
561	W H Q Q E M A N M K G A I G E R D D V Y Q R V R D Y ATGGCACCAACAAGAGATGGCAAATATGAAGGGTGCTATTGGTGAGAGGGATGATGTTTATCAAAGGGTCAGAGACTACC	640
641 721 801 881 961 1041	P S Q M P F A F R V Q P M Q P N L H E R I * S S T K K CATCCCAGATGCCTTTGCCTTCCGGGTGCAG <u>CCAATGCAGCCAAATCTTCA</u> TGAGAGGATTTAGAGCTCTACCAAAAA CAAAAGCATTTTAGCTAAGCTTTGCTGCTTGAAATGCAAGTTCTTAAGATTTCAGAAGTCTCCTTTGAGTAAAAGCCAT AATAAAGCAAACCTTGATGGTGCTATGAGACTATATATTGGATAGCTATATCTATC	720 800 880 960 1040

11) CaSEP1/2 (Coffea arabica SEPALLATA1/2)

AAACTCCAAATGTACAGGGGGTCCCAAGGACATGAAACCCATCAAACCAAAGTAGAAACACAGAAAAAGGGACTCCAACCCTAAAGCTTT CCCAGATTCCCATTTCTTCCCCCTTTTCTCTAACTACAACATAAATACATTACAATTTAGGGTTTGTATATATCTTCCTTGTTTC TTGCATAGTTTTGGTTGCAATTTGTGAGAAGTATAAAGGATCTAAGACTAACCCAGAAGAAGAAGGTTATTGTTGAGTAAAGATATATACTTT CACAGACATAGATATAAACAAGGGCTTGTGACTGCTGAGAAGTGTTGACCATGGGTAGGGGGGAGAGTGGAGCTGAAGAGAATAGAGAACAAGAT AAATAGGCAGGTTACGTTTGCTAAGAGGAGGAATGGACTGCTCAAGAAAGCTTATGAGCTCTCAGTTCTGTGTGATGCTGAGGTGGCCCTCATC ATCTTCTCTAACCGGGGCAAGCTTTATGAGTTCTGTAGCAGCTCCAACATGCTCAGGACTCTTGAAAGGTACCAAAGATGCAGTTATGGAGCAG TAGAAGTCAGCCACTCAGCCAAAGAGATTGAGCAAAGCAGCTACAAGGAGTACCTGAAGGCGAAAATATGAGTCACTACAACGATATCA AAGACACCTTCTTGGAGATGACTTGGGACCGCTAAATATAAATGATCTCGAACATCTTGAACATCAACTAGAATCATCCTTGAAGCTTGTAAGG TCCACAAGGACCaAAGTTATGCTGGATCAGCTTTCTGATCTCCAAACAAAGGAGAAGTTGTGGCTTGAGGCTAaTAAGGCATTaGAAAGGAaGC ${\tt ttgq} {\tt a} {\tt a} {\tt s} {\tt ttat} {\tt c} {\tt a} {\tt c} {\tt c} {\tt a} {\tt c} {\tt a} {\tt c} {\tt c$ TTCAcaGGGGTTTTTCcAGCCACTGGqATGCAGTTCAGGCTTGCAAATTGGGTACAATCCTGCAAGTTCAAGCCACATAACTGCAGTGACTAAT TGAAGATATATGCTGATGGTTTCCCCCTTGATACGTAAGCCTGTTATACAATTATGCAGCTACTCAATGATTAAAACGATAGTATTTGTCAGTCC ${\tt AGTCCCACAAGAACGAATTTAAGTACGAGATCAGGTAGATCATGCTACCTGGTTTTAGTGTTATAAGAAGGTGTATTTGACATCAAGTGTATTG}$ GCAAACAAACTTTATTTTGCAGAGTTATTTGCGTAAAATATGGAGTACATTTCGGTGCACTTTCAGCAATGGCTTCTGTTTTCCTTCAGTTTTA АААААААААААААААААААААААААААААААААА

Reads: CA00-XX-FB2-032-F12-RF.F CA00-XX-FB2-058-F07-EQ.F (clone) CA00-XX-FB4-013-F01-AB.F CA00-XX-FR1-029-F08-BF.F

>Proteína

MGRGRVELKRIENKINRQVTFAKRRNGLLKKAYELSVLCDAEVALIIFSNRGKLYEFCSSSNMLRTLERYQRCSYGAVEVSHSAKEIEQSSYKE YLKLKGKYESLQRYQRHLLGDDLGPLNINDLEHLEHQLESSLKLVRSTRTKVMLDQLSDLQTKEKLWLEANKALERKLGRYIYAENSPSSSWGD GGEQSITYSQQHSQYSQGFFQPLGCSSGLQIGYNPASSSHITAVTNAQNVSGLVPGWML-

1	AAACTCCAAATGTACAGGGGGTCCAAGGACATGAAACCCATCAAACCAAAGCAAAGTAGAAACACAGAAAAAGGGACTCC	80
81	AACCCTAAAGCTTTTGATCCCTCTTGTCTTTCCATCCAGATGCTAATCCTGAGGTTCAAGCTCAAACCCCCAATCCTGAAG	160
161	AGAAGTACAAGGAAAGAAGAAGAAGCAATCCCCAGATTCCCCATTTCTCCCCTTTTCTCTCTAACTACAAACATAAAT	240
241	ACATTACTACAATTTAGGGTTTGTATATATCTTCCTTGTTTCTTGCATAGTTTTGGTTGCAATTTGTGAGAAGTATAAAG	320
321	GATCTAAGACTAACCCAGAAGAAGAAAGGTTATTGTTGAGTAAAGATATATACTTTCACAGACATAGATATAAACAAGGG	400
	ACDC*EVLT <mark>MGRGRVELKRIENKINRQ</mark>	
401		480

481	<mark>V T F A K R R N G L L K K A Y E L S V L C D A E V A L</mark> GTTACGTTTGCTAAGAGGAGGAATGGACTGCTCAAGAAAGCTTATGAGCTCTAGTTCTGTGTGATGCTGAGG <u>TGGCCCT</u>	560
561	<mark>I I F S N R G K L Y E F C S S S N M L R T L E R Y Q</mark> <u>CATCATCTTCTCT</u> AACCGGGGCAAGCTTTATGAGTTCTGTAGCAGCTCCAACATGCTCAGGACTCTTGAAAGGTACCAAA	640
641	R C S Y G A V E V S H S A K E I E Q S S Y K E Y L K L GATGCAGTTATGGAGCAGTAGAAGTCAGCCACTCAGCCAAAGGAGTTGAGCAAAGCAGCTACAAGGAGTACCTGAAGCTG	720
721	K G K Y E S L <mark>Q R Y Q R H L L G D D L G P L N I N D L</mark> AAAGGCAAATATGAGTCACTACAACGATATCAAAGACACCTTCTTGGAGATGACTTGGGACCGCTAAATATAAATGATCT	800
801	<mark>E H L E H Q L E S S L K L V R S T R T K V M L D Q L</mark> CGAACATCTTGAACATCAACTAGAATCATCCTTGAAGGTTCAAGGTCCACAAGGACCAAAGTTATGCTGGATCAGCTTT	880
881	<mark>S D L Q T K E K L W L E A N K A L E R K L </mark> G R Y I Y A CTGATCTCCAAACAAAGGAGAGTTGTGGCTTGAGGCTAATAAG <u>GCATTAGAAAGGAAGCTTGGAAGA</u> TATATTTATGCT	960
961	E N S P S S S W G D G G E Q S I T Y S Q Q H S Q Y S Q GAGAATTCACCTTCATCATCGTGGGGGAGATGGTGGTGGGCAGAGCATAACATACAGCCAACAGCATTCTCAATATTCACA	1040
1041	G F F Q P L G C S S G L Q I G Y N P A S S S H I T A GGGGTTTTTCCAGCCACTGGGATGCAGTTCAGGCTTGCAAATTGGGTACAATCCTGCAAGTTCAAGCCACATAACTGCAG	1120
1121 1201 1281 1361 1441 1521	V T N A Q N V S G L V P G W M L * Q I N L Q I N I P T TGACTAATGCTCAAAATGTCAGTGGATTAGTTCCGGGGCTGGATGCTTTGACAAATTAATCTCCAGGATCAATATTCCAACC CCGGTCCACTTTCTTCTAAAGTGAAGATATATGCTGATGGTTTCCCCTTGATACGTAAGCCTGTTATACAATTATGCAG CTACTCAATGATTAAAACGATAGTATTTGTCAGTCCAGTCCAGTCCACAGGAACGAATTTAAGTACGAGATCAGGTAGATCATG CTACCTGGTTTTAGTGTTATAAGAAGGTGTATTTGACATCAAGTGTATTGGCAAACAAA	1200 1280 1360 1440 1520 1600
TOOT	ААААААААААААААААААААААААААААААААААА	

12) CaAG (Coffea arabica AGAMOUS)

XXX Splicing alternativo:

Reads:

CA00-XX-FB1-032-A11-EP.F CA00-XX-FB1-032-A12-EP.F CA00-XX-FB3-000-D09-EZ.F CA00-XX-FR1-032-B07-AC.F CA00-XX-FR1-032-E08-AC.F (clone)

>Proteína

MSCQSDPSRETSPQRKLGRGKIEIKRIENTTNRQVTFCKRRNGLLKKAYELSVLCDAEVALIVFSNRGRLYEYANNSVKETIKRYKTVNSDSAN TGSISEANAQHYQQEASKLRAQISNLQNSNRNMLGESLGSLNLRELKNIESKVERGISRVRSKKNELLFAEIEFMQKREVDLHNNNQYLRSKIA ETERAQHDMNLVPGSSDYELVSAQPFDARTFLQVNGLQLNNHYPRQEQRPLQLV-

Mapa do cDNA:

81	<mark>G K I E I K R I E N T T N R Q V T F C K R R N G L L K</mark> GGGAAAATTGAGATCAAGAGGATTGAAAACACCACAAATCGTCAGGTCACCTTCTGCAAGCGCCGCAATGGTTTGCTGAA	160
161	KAYELSVLCDAEVALIVFSNRGRLYE GAAAGCCTATGAATTATCTGTTCTGTGATGGTGGCGTTGTGGCGTCTTTATGAGT	240
241	<mark>Y A</mark> N N S V K E T I K R Y K T V N <mark>S D S A N T G S I S</mark> ATGCAAATAACAGTGTCAAAGAGACAATTAAGAGGTACAAAACGGTAAATTCAGATTCTGCGAACACCGGCTCTATTTCA	320
321	<mark>E A N A Q H Y Q Q E A S K L R A Q I S N L Q N S N R N</mark> GAAGCCAATGCTCAGCACTACCAACAGGAAGCATCGAAACTGCGTGCACAGATCAGTAATTTGCAGAACTCAAACAGGAA	400
401	M L G E S L G S L N L R E L K N I E S K V E R G I S CATGCTTGGTGAATCTTAGGGTCGTTGAATCTGAGGGGGGCTCAAGAATATAGAAAGCAAGGTTGAGAGAGGCATTAGCA	480
481	<mark>R V R S K K N E L L F A E I E F M Q K R E V D L H N N</mark> GAGTCCGATCCAAAAAGAATGAACTTTTGTTCGCGGAAATTGAGTTCATGCAGAAGAGGGAGG	560
561	NQYLRSKIAE TERAQHDMNLVPGSSDY AATCAATACCTTCGATGAAAGATGGAAAGAGGCCCAGCATGACATGGACTTGGTGCCTGGGAGCTCTGACTA	640
641	E L V S A Q P F D A R T F L Q V N G L Q L N N H Y P TGAACTAGTGTCAGCTCAGCCATTTGATGCTAGGACTTTCCTCCAAGTTAATGGGCTGCAATTAAATAATCATTACCCTC	720
721 801 881 961 1041	R Q E Q R P L Q L V * Y C L M R L R F L H L S F L V W GCCAGGAACAGAGGCCTCTTCAACTAGTCTGATATTGCTTAATGAGGTTGAGATTTCTCCATCTCTCATTTCTTGTTGG AAGTTCTGTCGCTAAGCTTGCCAGCAAGGACTAAATTCTTCAAGAAATCTGGGGGATTGATCG <u>GTGTTGGTGGAGGCGAAG</u> CTTCATTCTTAAAATAAGTATGCTGTTGGACAGAATATATAGTCCATTTTCTTCCAATTTCTATTGATAACATCTTTCTT	800 880 960 1040 1120
1121	AAA 1123	

13) CaAP1 (Coffea arabica APETALA1)

Reads:

CA00-XX-FR1-058-B11-QH.F (clone)

>Proteína

MGRGKVQLKRIENKINRQVTFSKRRAGLLKKAHEISVLCDAEVALIVFSHKGKLFEYSSDSSMENILERYERYSYAERRLVANDLESEGDWTLE YTKLKAKIELLQRNHRHYMGEDLDAMSSKDLQNLEHQLDTALKQIRTRKNQLMYESISELQRKEKAIQQQNSMLAKKIKEKEKLMAQQAQWEQQ NQGPSSTPYLIPEPLPPCINVSGNYEEETQEARRNDLELTLDSLFPCHLGCFTA-

1	GTGATTTTAGAATCTAGGGGCATGGGATGATAAATTGGAGTTAGAGGGTTTCCTTTTTTAGATCCTGTGTGAAGAGGGTC	80
81	R L F S G E K K R E R G E I <mark>M G R G K V Q L K R I E N</mark> GTTTGTTTCTGGGGAGAAAAGAAAAGAGAGAGAAGAAATCATGGGGAAAGGTACAGTTGAAGAAAAAGAAAAAGAAA	160
161	<mark>K I N R Q V T F S K R R A G L L K K A H E I S V L C D</mark> AAGATTAACAGGCAAGTGACATTTTCAAAGAGAAGAGCTGGATTGTTGAAGAAAGCTCATGAAATATCAGTACTTTGTGA	240
241	<mark>A E V A L I V F S H K G K L F E Y S S D</mark> S S M E N I TGCTGAAGTGGCTTTGATCGTCTTCTCTCACAAAGGGAAGCTCTTTGAATACTCTTCTGATTCTAGCATGGAGAATATCC	320
321	L E R Y E R Y S Y A E R R L V A N D L E S E G D <mark>W T L</mark> TGGAACGATATGAAAGATATTCATATGCAGAAAGGCGATTAGTTGCAAATGATCTAGAATCCGAGGGAGATTGGACTCTG	400

	EYTKLKAKIELLQRNHRHYMGEDLDAM	
401	GAATACACCAAACTCAAGGCCAAGATTGAGCTGTTGCAAAGAAATCATAGGCACTACATGGGTGAGGACCTTGATGCA <u>AT</u>	480
481	<mark>S S K D L Q N L E H Q L D T A L K Q I R T R K N Q L</mark> GAGCTCGAAAGACCTCCAGAATTTGGAGCACCAGCTTGACACGGCTCTGAAACAAATCCGAACTAGAAAGAA	560
561	<mark>M Y E S I S E L Q R K E K A I Q Q N S M L A K</mark> K I K TGTACGAGTCAATCTCTGAGCTGCAGAGAAAGGAAAAAGCAATTCAACAGCAAACAGCATGCTCGCAAAGAAGATAAAG	640
641	E K E K L M A Q Q A Q W E Q Q N Q G P S S T P Y L I P GAGAAAGAGAAGTTAATGGCACAACAGGCACAGTGGGAGCAGCAAAATCAAGGCCCGAGTTCAACTCCATACCTGATTCC	720
721	E P L P P C I N V S G N Y E E E T Q E A R R N D L E AGAACCACTCCCTCCTTGCATGTCAGTGGCAACTATGAAGAAGAAACCCAAGAAGCGAGAAGGAATGACCTTGAGC	800
801 881 961 1041 1121	L T L D S L F P C H L G C F T A * S K R K I G C Y T C TTACTCTAGACTCGCTGTTTCCCTGCCATCTTGGATGCTTTACTGCCTGATCAAAGCGCAAAATTGGGTGCTATACTTGC TAGTACTTGTAACTGCATACGTTGACACC $\underline{GATAAAGGCGGAGAAACCA}$ AAATTGCACCCAACTTGGAAATTTCTAATAT GTTTGACTATCTGCTGGTGTTATCTGCGCTAAGTGTATTTTGTGAACCCGATCTGGAAACTTTTCAGCTGTCTCTAAAG CTGTAGTTCTTGAGCATTTTTAGAATGATGGAAAAAGAAAAAAAA	880 960 1040 1120

14) CaTM6 (Coffea arabica TM6)

XXX Splicing alternativo: 1 - poli-A 2 -

Reads:

CA00-XX-FB1-021-C05-UT.F CA00-XX-FB1-040-F06-JE.F CA00-XX-FR2-022-F12-CB.F (clone) CA00-XX-LV8-072-G04-AG. CA00-XX-SH2-043-E08-EM.F

>Proteína

MGRGKIEIRKIENSTNRQVTYSKRRNGIFKKAHELSVLCDAKVSLIMLSDTKKFHEYTSPSITTKKVMDDYQSAVGVDLWSTHYEKMQENLRRL KETNNKLRRDIRQRMGEDLNDLNWDNMCRLQEKIVDSLAIIRHRKYHVIKNQTDTYKKKVRNLKERHGNLLYDLEARSCEDPKYGIVDNARDYN SALALADGGLSNLYALRLQSSHPNLQDLRLA-

```
Mapa do cDNA:
```

	Н	A	S	G	F	Q	ΕH	E I	4 (G I	R (G F	< I	F	I I	R	K	I	Е	Ν	S	Т	Ν	R	Q	
1	CCCAC	CGCG	TCC	GGA:	TTCC	CAGG	AGG	AGA	rgg	GTCO	GTG	GAAA	AGAT	AGA	GAT	CAG	GAA(GATI	rga(GAAC	CTCA	ACC	AAC	AGG	CAA	80
	V 1	ГY	S	K	R	R	Ν	G	Ι	F	K	K	A	H	E	L :	S I	Ι	5 (C E	A	K	V	S	L	
81	GTTAC	CTTA	CTC	TAA	GAGA	AGA	AAT	GGT	ATT	TTC?	AAG	AAA	GCCC	CATG	GAGC	TCA	GCGI	FTCI	TTT	GTGA	TGC	TAA	GGT	TTC.	ACT	160
	I	М	L :	s i	r c	K	K	F	Н	E	Y	Т	S	Р	S	I	Т	Т	K	K	V	М	DI	D	Y	
161	CATCA	ATGC	TCT	CCGI	ATAC	CAA	.GAA(GTT(CCA	IGA	GTA	TACI	FAGT	CCI	TCA	ATA	ACGI	ACTA	AAA	AAGO	STCA	TGG	ATG	ATT.	ATC	240
	Q S	А	V	G	V	D	L V	NĪ S	S 1	ΓH	H I	ΥE	E K	C M	1 Q	E	N	L	R	R	L	K	Е	Т	Ν	
241	AAAGO	CGCT	GTA	GGT	GTTG	GATC	TCTO	GGA	GCA	CGCA	ACTZ	ATGA	AGAA	AAT	GCA	AGA	GAAI	FTTÆ	AGA	AGA	TTG	AAG	GAG	ACC.	AAC	320
	NF	ΚL	R	R	D	Ι	R	0	R	М	G	Е	D	L	N	DI	LÌ	V V	V I) N	I M	C	R	L	0	
321	AATAA	AGCT	GAG	GAG	AGAC	CATC	AGG	CAA	AGA	ATG	GGT	GAAC	GACI	TGA	ATG	ATC	ГСА <i>Р</i>	ACTO	GGGZ	ACAA	TAT	GTG	CCG'	ГСТ	GCA	400
			_														_			_	_	_	_			
	E	K	1 '	V I) 5	5 L	A	1	T	R	Н	R	K	Y	Н	V	1	K	Ν	Q	Т	D	Т	Y.	K	
												~	~													

401	AGAGAAGATTGTTGATTCTCTTGCTATCATACGCCATAGAAAGTACCATGTGATCAAAAATCAAACTGATACCTACAAGA	480
481	K K V R N L K E R H G N L L Y D L E A R S C E D P K Y AGAAGGTGAGAAACCTGAAGGAAAGACATGGAAATCTTCTGTATGACTTGGAGGCAAGATCATGTGAAGATCCAAAGTAT	560
561	G I V D N A R D Y N S A L A L A D G G L S N L Y A L R GGGATTGTGGACAACGCCAGAGACTACAACTCGGCCCTTGCATTAGCTGATGGAGGGTTGTCCAACCTCTATGCTCTCCG	640
641	L Q S S H P N L Q D L R L A * T L K A P R T G D L S ATTGCAGTCGAGCCACCCCAATCTTCAGGACCTCCGACTTGCTTAAACTTTAAAAGCTCCAAGAACTGGGGATTTATCTA	720
721	N L C S Y L I L S A K N * W M C * S S Y E K E C * K * ATTTATGTTCTTATCTAATTTTATCCGCAAAGAACTAGTGGA <u>TGTGCTAGTCCTCGTATGAGAA</u> AGAATGTTAGAAGTAA	800
801	F K L L K L K N C I N Y W D E V F A H A I S T S M L * TTTAAGTTGTTGAAACTCAAAAACTGCATTAACTATTGGGACGAGGTTTTTGCTC <mark>A</mark> TGCAATATCTACTTCGATGCTGTA	880
881	C D F * F P L K K K K K K K K atgcgatttttaatttcctttaaaaa <mark>a</mark> aaaaaaaaaaaaaaaa	

15) CaFUL (Coffea arabica FRUITFULL)

Reads: CA00-XX-FR4-015-C03-AC.F (clone) CA00-XX-FR4-038-H07-JM.F

>Proteína

MGRGRVQLKRIENKINRQVTFSKRRSGLLKKAHEISVLCDAEVALIVFSTKGKLFEYATDSCMERILERYERYAYAERQLGGAEIESQGCWSLE HAKLKARIEVLQRNQRHYMGEDLDNLSLRELQNLEHQLDTALKHIRSRKNQLMFESIAELQKKDKALQEQNNVLAKKVKDKEKEQAQQPQWERQ NRHDLNSSSLVRSQPINSLSISETYHRGGDNEAEGTQSSQTNAVMHPWMLRQMN-

1	CTCACAAACTCGTCTTTTTCATCTTGCACGAAAAGTCATAGAGACATGATTATTGGTATATAGCGAGAATATTAATCGG	80
81	L I W I F I F * F N S T R E R R R <mark>M G R G R V Q L K</mark> TTTAATTTGGATTTTCATCTTTTGATTTAACAGCACGAGAGAAGAAGAAGAATGGGGAGGGGGGGG	160
161	R I E N K I N R Q V T F S K R R S G L L K K A H E I S GAATCGAGAACAAGATTAACAGGCAAGTGACTTTCTCCAAAAAGGCGATCTGGGCTGCTCAAGAAAGCTCATGAGATCTCC	240
241	VLCDAEVALIVFSTKGKLFEYATDSCM GTGCTTTGCGATGCTGAGGTTGCTTTGAGTATGCTACTGATTCATGCAT	320
321	<mark>E R I L E R Y</mark> E R Y A <mark>Y A E R Q L G G A E I E S Q G</mark> GGAAAGGATCCTTGAGAGGTATGAAAGGTATGCCTATGCAGAAAGACAGCTTGGGGGTGCAGAAATCGAGTCACAGG <u>GCT</u>	400
401	C W S L E H A K L K A R I E V L Q R N Q R H Y M G E D GCTGGAGTCTGGAACATGCAAAACTCAAGGCCAGGATTGAAGTTCTACAAAGAAACCAAAGGCACTATATGGGAGAAGAC	480
481	<mark>L D N L S L R E L Q N L E H Q L D T A L K H I R S R K</mark> CTTGACAATTTAAGTCTCAGAGAGCTTCAGAATTTGGAACATCAGCTTGATACAGCTCTTAAACACATCAGGTCAAGAAA	560
561	NQLMFESIAELQKKDKALQEQNNVLA GAACCAGCTCATGTTTGAATCAATTGCTGAGCTGCAGAAGGAGAAAGGCACTGCAGGAGCAAAACAACGTGCTTGCAA	640

	<mark>K K V K D</mark> K E K E Q A Q Q P Q W E R Q N R H D L N S S	
641	AAAAGGTGAAAGACAAGGAGAAAGAACAAGCCCAACAGCCACAGTGGGAACGACAGAACCGTCACGACCTGAATTCATCA	720
	5. V R S O P T N S I. S T S F T V H R C C D N F & F C	
7.01		000
/21	TCATTGGTTAGGTCACCAACCAATCAACTCCTTGAGCATTAGTGAAACGTACCATCGTGGTGGAGATAATGAAGCTGAAGG	800
	T Q S S Q T N A V M H P W M L R Q M N * I D R R L R	
801	AACTCAGAGCAGCCAAACCAATGCAGTCATGCACCCATGGATGCTCCGCCAGATGAATTGAATTGATAGAAGATTACGGC	880
881	TGGTGGTTTTTGGCCACAAACTACTAATGATAGGCTTATAAATTGTTGAAGAACAATGTTTAAAAGCGGTGCCCACCATCC	960
961	ATTCCCCGAGAAAAGGTAATACTAATGATGTAAACGTGTGCTGTAAAAGGCTGTGGCTGTTACCCAGAAAAAAAGCTCC	1040
1041	ACTTTAATTCATATGTTACGGTATGCATGAATGAGGATTGAAATACTATAAAGTAACCCTACC <mark>AAAAAAAAAA</mark>	1120
1121	ATGTAAATATCTGTGCATACTAGATTCAAAAAAGGAAGCTCATCTTTGTGAGGAAGCAAGACCTATACATGTTGGTTCA	1200
1201	GTTTCAAAAAAAAAAAAAAA 1221	

16) CaAGL42 (Coffea arabica AGAMOUS-like42) - sem clone e incompleta

Reads:

CA00-XX-IC1-022-D07-EC.F

>Proteína (Incompleta)

RKMVRGKIQMRRIENATSRQVTFSKRRNGLLKKAYELSVLCDAEVALIIFSQKGKLYEFSSSNMQKTIDKYRGCVKEDQRSDQDIEKYVQELKL EAINMANTIEFLEASQRKLLGQDLGSSSLRELPQIDSQLERSLKRNVRARRTQLFK

Mapa do cDNA:

81 AAACCTTATTGTCTTCTGTGGTCTGAAAATAAAGAGCTTGCTACATCAACCATCAGAAGATCCTTCATGTCCTT 161 CTTCTTCTTCTTCTTCTTCTCGCCTATTTAGAGTCCTTGGCTTATTATTTAGTTATTCGCAAAG	TTTCTT 160 CGACTA 240
161 CTTCTTCTTCTTCTTCTTCTTCAGCTTGCCTATTTAGAGTCCTTGGCTTATTATTATTTAGTTATTCTGCAAAG	CGACTA 240
241 GCAT <u>TGACCCTCCCAACTCATTTC</u> TAAACTCTTTCCAGAAAATTACTTTTCCATAATTTTCTAGGCTTCAATTC	TATGCT 320
V G D L G S H T T C R K <mark>M V R G K I Q M R R I E</mark>	N A
321 TGTGGGAGATTTGGGCTCTCATACAACTTGCAGAAAAATGGTGAGAGGGAAGATTCAGATGAGAAGGATTGAGA	ATGCAA 400
T S R Q V T F S K R R N G L L K K A Y E L S V L C	D A
401 CGAGCAGGCAGGTGACCTTCTCAAAGAGGAGAAATGGGCTTCTGAAGAAGGCGTACGAGCTATCAGTTCTGTGC	GATGCT 480
E V A L I I F S Q K G K L Y E F S S S N M Q K T I	D K
481 GAAGTTGCTTTGATAATTTTCTCACAAAAAGGGAAGCTCTATGAGTTCTCAAGCTCCAACATGCAAAAGACAAT	CGATAA 560
<mark>Y</mark> R G C V K E D Q R S D Q D I <mark>E K Y V Q E L K L</mark>	E A
561 ATATCGTGGCTGTGTGA <u>AGGAAGATCAAAGAAGCGACC</u> AAGACATTGAGAAATATGT <u>ACAGGAGCTGAAGCTGG</u>	AAGCTA 640
INMANTIEFLEASQRKLLGQDLGSS	S L
641 TAAACATGGCAAATACGATAGAATTCCTTGAAGCTTCTCAACGGAAGCTTTTGGGGCAAGATCTAGGATCAAGT	TCACTA 720
R E L P Q I D S Q L E R S L K R N V R A R R T Q L	, F K
721 AGAGAACTTCCACAGATTGACAGCCAGCTCGAGAGAAGTCTTAAACGGAACGTCCGGGCAAGGAGAACTCAGCT	ATTCAA 800

801 GGA 803

17) CaSVP-2 (Coffea arabica SHORT VEGETATIVE PHASE-2)

Reads: CA00-XX-CS1-050-A10-EP.F CA00-XX-CS1-112-B05-EQ.F CA00-XX-LV4-041-H02-AB.F CA00-XX-LV5-006-D04-RF.F CA00-XX-LV5-034-E10-CB.F (clone) CA00-XX-LV5-063-A03-AC.F CA00-XX-RX1-042-B05-EB.F CA00-XX-RX1-072-D05-EB.F CA00-XX-RX1-083-C01-EB.F

>Proteína

MAREKIKIKKIDNITARQVTFSKRRRGLFKKAEELAVLCDADVALIIFSATGKLFEFASSSMSDILGKYKLHSSNLEKTEQPSLELQLENSCHV RLSKEVADRTHQLRQMKGEDLQGLKIEELQQLEKVLEAGLTRVLQTKGERIMNEINALQKKGAELFEENKQLKQKMAMLYEGKRPVIPDLDKDM LIEEGQSSESITNVCSCNSGPPPEDDCSDTSLKLGLPFN-

Mapa do cDNA:

1 81 161	GGAAAACTGAGGATAAAAGAGCGATTGGTTTTGCTATAAAAGGGAAAGTGGACTTTCCTTTCTTGACATGGGGCTTCTGT TCCCAATTCTTCCCTCCCTCTCTCAATTGGTGGCTTCTTTTGATATTTTCGATTCCATATAGGATATATTTTTTTT	80 160 240
241	H E T K N I S C S S S P S H S S T Q * G D R K E K <mark>M A</mark> CACGAAACAAAGAACATCTCCTGCAGCTCCTCACCCCAGTACCCAATAAGGCGATCGAAAAGAAAAAATGGC	320
321	REKIKIKKIDNITARQVTFSKRRGL GAGAGAAGATCAAAATCAAAAATCGATAACATCACGGCGAGGCAGGTGACCTTCTCCAAGAGGAGACGAGGGCTTT	400
401	FKKAEELAVLCDADVALIIFSATGKLF TCAAGAAAGCTGAAGAGCTTGCTG <u>TTTTGTGCGATGCTGATGTT</u> GCCCTCATAATCTTCTCAGCCACTGGCAAGCTCTTC	480
481	<mark>E F A S S S M S D I L G K Y</mark> K L H S S N L E K T E Q P GAGTTCGCTAGCTCAAGCATGAGTGATATCCTTGGAAAGTACAAGTTACATTCAAGTAACCTCGAGAAGACTGAGCAACC	560
561	S L E L Q L E N S C H V <mark>R L S K E V A D R T H Q L R</mark> ATCGCTTGAGCTTCAGCTAGAGAATAGTTGCCATGTTAGATTGAGCAAGGAAGTTGCTGACAGGACTCATCAGCTAAGGC	640
641	<mark>Q M K G E D L Q G L K I E E L Q Q L E K V L E A G L T</mark> AAATGAAGGGTGAGGACCTTCAAGGGTTAAAGATAGAGGAGCTGCAGCAGTTAGAAAAAGTGCTGGAAGCAGGACTAACC	720
721	<mark>R V L Q T K G E R I M N E I N A L Q K K G A E L F E E</mark> CGTGTGCTTCAAACGAAGGGTGAACGCATCATGAATGAGATCAATGCCCTTCAAAAGAAGGGTGCAGAACTGTTTGAAGA	800
801	NKQLKQK MAMLYEG KRPVIPDLDKDM AAATAAGCAACTGAAGCAAAAAATGGCAATGCTTTATGAAGGG <u>AAGAGGCCAGTGATTCCAGA</u> CTTGGATAAAGATATGC	880
881	L I E E G Q S S E S I T N V C S C N S G P P E D D C TGATAGAAGAAGGCCAATCTTCAGAGTCCATTACCAATGTTTGCAGCTGCAACAGTGGCCCTCCTCCGGAGGATGACTGC	960
	S	
961	TCAGATACATCACTCAAACTAGGGCTACCCTTTAACTAGCCGACGGTGGACAGCTTATGGGGACCTACCT	1040
1041	CTAGTAACATGGGATCAAGAAATGGAGCTCTCTCCATCTATATGTCTATATGTTATCCAAGATAGTGGTAAAAGTCTATG	1120
1121	GTAGTATTACTCCTGTCATAAAATTAAGAGTCTGCTGTGTATGGCTGGC	1200
1201	AAAAAGCTCGTGTTGCATCCTGCTTCCAACCAGAACTCAAGATTTGTACTCTCTTGTATGAACTTGCTTCCGGGCTTC	1280
1281	AGCTGAAAAGTTAATCCTAATAATATGAAATGCTGCTGGTAGAACTATTTATCTAAAAAAAA	1360
⊥J6⊥	АААААААААААААА 13/5	

18) CaAGL24 (Coffea arabica AGAMOUS-like24)

AGGGCTTTTCAAGAAAGCTCAGGAGCTCTCCACACTTTGTGATGCTGAGATAGCCCTCATAGTCTTCTCAGCAACAGGCAAGCTCTTCGAGTAT TGTAGCTCAAGTATGCAAGTTATTGAAAGGCACCGCCTATGTTCCGAGGACATTGGGAGGCAAGACAAGCATCCACCTCATCTCACGCAGC GGGAGAATCATACCCATGCCATGCTTGCTGAGGAAATCAAGGAGAAAACCGCAGAACTGAGGCATCTCAAAGGTGAAGAGCTAGTAGGACTGAG ATTCTCAAGAAAAAGGAAGCATTATTAAAGGAAGAGAATGCAAAGTTGAGACAAAAAGTAGCAGGCACATCGGAAGACGAAACACCTCTGCTGG AACAAGGGATTTCATCCGAGTCAGTGACCCGCTTAAGCGACCAAGCCGGAAGTTGTCCTCAGGACCTCAACAATTCGGACACATTTCTCCAGCT GGGGTATGTTTGTTTCAAGTCAAACATGGATCGGATAATTTGTTACAAGAATAAAACGAATTACATATATGGAATATAATATGTGAAATCTATC ${\tt CAACATTTGGATCTTATATCTGGACGACGATGCTGCATATCAGTTGAATGGGGGGGCATCTCATTCTGTTTGGTTTTGCCACAACAGATGGCTAGT$ ATTTAAGCAAGAAGAACAAAACGAAACTAAAGTAAGATTTCATTTTCAGAATGCATGATTCAAGGCCATGTGATCGTCTGGCAATGACGTGAAA TAACTTGGGGTTGAATTTTTGCAGCTTACCATGTCCCCAACTGAATTGCAGGGAGGATGAGATTGCGGTCGCAGAAGATCGAATATATAATTTGA ${\tt TGTGTATATATATATATAGGTACAGATGAGGGATGTAATTTATGTTGAATTGCTTTGCAGCTTAAATAGTTTTCTGACAGCGTGATTAATGA$

Reads:

CA00-XX-LV8-006-F01-BF.F (clone)

>Proteína

MVRQRIQIKRIDNLTARQVTFSKRRRGLFKKAQELSTLCDAEIALIVFSATGKLFEYCSSSMMQVIERHRLCSEDIGRQDKHPPHLTQRENHTH AMLAEEIKEKTAELRHLKGEELVGLSMEDLVKLEKLVEAGLSRIAKTKGDKFMKEIGILKKKEALLKEENAKLRQKVAGTSEDETPLLEQGISS ESVTRLSDQAGSCPQDLNNSDTFLQLGYVCFKSNMDRIICYKNKTNYIYGI-

1 81 161	ACCGGAATCCTCTCAAGGGTCTGTCCAGTTGAGTTAAGCTTCAGACATCAGATTGACAAGGAAAAAATAGAAAGTAAAAG GTAGTACTCCATTCCTCCTCATATTTAGTTTTACTTGTACTAGTTCTCTCTC	80 160 240
241	I L I G I L S A G L T S D L S S E S R K <mark>M V R Q R I Q</mark> TCTTAATTGGCATTCTTTCGCAGGGTTGACGTCCGATCTTTCA <u>TCCGAGTCGAGAAAAATGGT</u> CAGGCAAAGAATTCAG	320
321	<mark>I K R I D N L T A R Q V T F S K R R R G L F K K A Q E</mark> ATCAAGAGGATTGATAACTTGACAGCTAGACAAGTGACTTTTCAAGAAGAGAGAG	400
401	<mark>L S T L C D A E I A L I V F S A T G K L F E Y C S S</mark> GCTCTCCACACTTTGTGATGCTGAGATAGCCCTCATAGTCTTCTCAGCAACAGGCAAGCTCTTCGAGTATTGTAGCTCAA	480
481	<mark>S M M Q V I E R</mark> H R L C S E D I G R Q D K H P P H L <mark>T</mark> GTATGATGCAAGTTATTGAAAGGCACCGCCTATGTTCCGAGGACATTGGGAGGCAAGACAAGCATCCACCTCATCTCACG	560
561	<mark>Q R E N H T H A M L A E E I K E K T A E L R H L K G E</mark> CAGCGGGAGAATCATACCCATGCCATGCTTGCTGAGGAAATCAAGGAGAAAACCGCAGAACTGAGGCATCTCAAAGGTGA	640
641	<mark>E L V G L S M E D L V K L E K L V E A G L S R I A K</mark> AGAGCTAGTAGGAC <u>TGAGCATGGAAGACTTGGT</u> TAAACTCGAAAAGTTGGTTGAAGCAGGGTTAAGTCGTATTGCCAAAA	720
721	<mark>TK</mark> G DKFMKEIGILKKKEALLKEENAKL CCAAGGGTGACAAATTCATGAAAGAAATTGGTATTCTCAAGAAAAAGGAAGCATTATTAAAGGAAGAAAGGAAAGTTG	800
801	R Q K V A G T S E D E T P L L E Q G I S S E S V T R L AGACAAAAAGTAGCAGGCACATCGGAAGACGAAACACCTCTGCTGGAACAAGGGATTTCATCCGAGTCAGTGACCCGCTT	880
881	S D Q A G S C P Q D L N N S D T F L Q L G Y V C F K AAGCGACCAAGCCGGAAGTTGTCCTCAGGACCTCAACAATTCGGACACATTCTCCAGCTGGGGTATGTTTGTT	960
961 1041 1121 1201 1281 1361 1441 1521 1681 1761 1841	S N M D K I I C Y K N K T N Y I Y G I * Y V K S I Q H CAAACATGGATCGGATAATTTGTTACAAGAATAAAACGAATAACATATATGGAATATAATATGGAAATCTATCCAACAT TTGGATCTTATATCTGGACGACATGCTGCATATCAGTTGAATGGGGGGCATCTCATTCTGGTTTGGCACAACAA GGCTAGTATTAAGCAAGAAGAACAAAACGAAACG	1040 1120 1280 1360 1440 1520 1600 1680 1760 1840 1920
1921	АААА 1924	1920

19) CaFLC (Coffea arabica FLOWERING LOCUS C)

Reads:

CA00-XX-RX1-029-B04-EB.F (clone) CA00-XX-RX1-048-B05-EB.F CA00-XX-SH2-052-D10-EM.F

>Proteína

MGRRKVEIKKIEDKNSRQVTFSKRRSGLMKKAKELSVLCDVDVAVLIFSGRGKLYDFCSTNSLAKILQRYRNYAEAEDGSARISGVEKRNPEGR NVVTIRKLLEKVERDLEEPDVDHLNLSELVQLEEQLEDALIQTRSRKTRLLMESITSLSEVEKMLREENKLLQNKVAAGTSNEKRNDLILEFGD LTHVGMISGQRQAMLELL-

Mapa uu CDNA.	Mapa	do	cDNA:
---------------	------	----	-------

1 81 161	GATCCCCAGGCCCTACAGTAAAATAATAATTCTGGCGTAGGAATCCTTCGACTGCCGTCTTTATGAAATTCTCATTCCTT CTCCTCTGTACATATACCACTCACTCACAGCTAGGTAGAAACGAATACAGGTTCAAATTCCGAATCTCGGAGAATTAGGG CAAAAACCAGAACCAACGATCTGTTTTTTCTTATCTCCCAGTTCGTACCTTTTTTTT	80 160 240
241	P T K E L R <mark>M G R R K V E I K K I E D K N S R Q V T</mark> ACCGACGAAGGAATTAAGGATGGGGCGGAGGAAGGTGGAGATTAAGAAAATCGAGGACAAGAACAGCAAGGCAAGTCACGT	320
321	<mark>F S K R R S G L M K K A K E L S V L C D V D V A V L I</mark> TTTCCAAGCGGAGAAGCGGACTGATGAAGAAAGCCAAGGAACTTTCCGTTCTCTGCGACGTGGATGTTGCTGTCCTCATC	400
401	<mark>FSGRGKLYDFCSTNSLAKILQRY</mark> RNYA TTCTCTGGTCGCGGCAAGCT <u>CTACGACTTCTGCAGCACCA</u> ACAGTTTGGCCAAGATCCTACAACGATATCGCAACTACGC	480
481	E A E D G S A R I S G V E K R N P E G R N V V T I R AGAAGCAGAAGACGGGTCTGCAAGAATTAGCGGCGTAGAGAAACGTAACCCTGAAGGCAGAAATGTCGTGACAATCAGAA	560
561	K L L E K V E R D L E E P D V D H L N L S E L V Q L E AGCTGCTGGAAAAAGTTGAAAGGGATCTAGAGGAGCCAGATGTTGACCACCTTAACCTGAGTGAACTAGTGCAATTGGAA	640
641	E Q L E D A L I Q T R S R K T R L L M E S I T S L S E GAACAACTTGAAGATGCACTCATTCAAACAAGATCTAGGAAGACACGATTACTGATGGAATCAATAACCAGTCTAAGTGA	720
721	V E K M L R E E N K L L Q N K V A A G T S N E K R N AGTGGAAAAGATGCTGAGGGAAGAAAAAAAGCTTCTGCAAAATAAGGTAGCTGCAGGTACATCCAATGAGAAGAGGAATG	800
801 881 961 1041	D L I L E F G D L T H V G M I S G Q R Q A M L E L L * ACTTG <u>ATCCTTGAATTTGGGGACCTT</u> ACACACGTTGGAATGATTTCTGGGCAGCGACAGGCTATGCTTGAACTACTTAA GAGTAGAAGATAGTGAGACAAGCCAGTTTGTGCCTTTCTGAGTGCTCCAAGGGAAAATATACCCGTTAAGGCAATGTGAT CCTACTATACAACTCTTCATAAGCAACTTTTTTATAACTTGGTTTTGCGTCTGACATCACTCCGCTTCATAACATGAGATT GAAGTGGGGGTACGTTCTCCCACTTGCATTGTTCTGTCAAAAAAAA	880 960 1040

20) CaSOC1 (Coffea arabica SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1) - sem clone e incompleta

Reads: CA00-XX-FB4-017-B06-MC.F CA00-XX-IA2-029-E06-EC.F CA00-XX-SI3-024-H04-EM.F

>Proteína

VTFSKRRNGLLKKAFELSVLCDAEVALIIFSPRGKLYEFGSSSMKEIIERYQKHAKDVRANNPSAEQNMQQLKQETASMVKKIELLEASKRKLL GEGLVSCTVEELQQLERQLERSVNCIRARKMQVFQEQIEKLKEKEKVLEAENDKLLEKCGAEPPQTSKENTEIVPCTESSEVSDVETGLFIGPP ERRNKLVLKN-

Mapa do cDNA:

1	<mark>V T F S K R R N G L L K K A F E L S V L C D A E V A L</mark> GTGACCTTCTCCAAGAGGAGAAATGGCCTGCTGAAAAAGGCTTTCGAGCTTCAGTTCTTGTGATGCTGAAGTTGCACT	80
81	<mark>I I F S P R G K L Y E F G S S</mark> S M K E I I E R Y Q K CATCATCTTCTCTCCCAGAGGCAAGCTCTATGAATTTGGTAGTTCAAGCATGAAGGAGATAATAGAACGCTATCAGAAGC	160
161	H A K D V R A <mark>N N P S A E Q N M Q Q L K Q E T A S M V</mark> ATGCAAAAGATGTTCGAGCTAACAACCCCTCAGCAGAACAGAATATGCAGCAACTGGAAACAGGAAACTGCAAGCATGGTG	240
241	<mark>K K I E L L E A S K R K L L G E G L V S C T V E E L Q</mark> AAGAAGATAGAGCTCCTTGAAGCTTCCAAAAGGAAACTTTTGGGGGAAGGCTTAGTTTCGTGCACGGTTGAGGAACTGCA	320
321	<mark>Q L E R Q L E R S V N C I R A R K M Q V F Q E Q I E</mark> GCAGTTGGAGCGCCAATTAGAACGAAGTGTGAACTGCATTCGAGCAAGAAA <mark>GATGCAAGTGTTCCAGGAACAAATTGAAA</mark>	400
401	<mark>K L K E K E K V L E A E N D K L L E</mark> K C G A E P P Q T <mark>AATTAAAAGAAAA</mark> GGAAAAAGTCCTGGAAGCTGAAAATGACAAGTTATTGGAGAAGTGTGGAGCAGAGCCTCCACAGACA	480
481	S K E N T E I V P C T E S S E V S D V E T G L F I G P TCAAAAGAGAACACAGAGATTGTGCCTTGTACAGAGAGTAGCGAAGTTTCAGATGTGGAGACTGGGCTATTTATT	560
561	PERRNKLVL <mark>KN*</mark> RPSQMKTLLCLALH ACCGGAGAGAAGAAATAAGCTAGTGTTGAAGAACTGA <mark>CGA</mark> CCTAGCCAGATGAAGACTCTTCTCTGC <mark>TTGGCTCTTCATC</mark>	640
641 721 801 881	L Q R A D C V S L * R L F Q C Q I * N S L * L L D C G TGCAGCGAGCTGACTGC <mark>GTATC</mark> ATTATGACGATTGTTTCAATGTCAAATATGAAATAGCCTATAGCTGCTAGACTGTGGA AAGCAAGAATCGGAGAGCATTATTACGCCTCAAGAAGTGATAAACGAACCTGTGATTGCATTATTGGTGTTAGTTTTATTA ATTTTGGCATCGTATTCCCAATGGAATACGATCAAGAAAAATCTTTCATAGGGTCTACCCTGAGGTATATTAATATATAT	720 800 880
XXX	Regiõs de spling alternativo	

21) CaAGL6 (Coffea arabica AGAMOUS-like6)

Reads: CA00-XX-FB2-025-G08-JM.F CA00-XX-FB2-030-A10-JF.F CA00-XX-FB2-031-H12-RF.F (clone)

>Proteína

MGRGRVELKRIENKINRQVTFSKRRNGLLKKAYELSVLCDAEVALIIFSSRGKLYEFGSSGITKTLERYQRCSLSPQENAAERETQSWYQEVSK LKAKYESLQRAQRHLLGEDLGPLNVKELQNLEKQLEGALLQARQRKTQLMIEQMEELRRKERQLGDLNKQLKIKVSLEMSSLEAAEGQGLIRGL PWLWSSSVPSGSSMFPMHTSHPSAAMDCDPEPVLQIGYHQYAPAEGPSAPRSMAIESNIIQGWAL-

161	T H A C H S I I S F F V E T F S R G K K N <mark>M G R G R V</mark> ACACACGCTTGCCATTCAATCATAAGTTTTTTGTAGAAACATTTTCAAGGGGAAAAAAAA	240
241	<mark>ELKRIENKINRQVTFSKRRNGLLKKA</mark> AGAGTTGAAGAGGATAGAGAACAAGATCAACCGTCAAGTGACTTTCTCTAAGAGAAGAAATGGTCTGCTCAAGAAAGCTT	320
321	Y E L S V L C D A E V A L I I F S S R G K L Y E F G S ATGAACTTTCTGTTCTTGTGATGCTGAAGTTGCTCTCATCATTTTTCAAGTCGCGGCAAGCTCTACGAGTTCGGCAGC	400
401	<mark>S G I T K T L E R Y</mark> Q R C S L S P Q E <mark>N A A E R E T Q</mark> TCTGGGATAACTAAAAACCCTTGAGCGATATCAACGTTGCAGCTTGAGTCCTCAG <u>GAGAATGCTGCTGAAAGAGA</u> AACACA	480
481	<mark>S W Y Q E V S K L K A K Y E S L Q R A Q R H L L G E</mark> GAGCTGGTACCAAGAGGTCTCAAAACTGAAGGCCAAGTACGAATCACTACAACGCGCTCAAAGGCACCTACTTGGAGAAG	560
561	<mark>DLGPLNVKELQNLEKQLEGALLQARQR</mark> ATCTTGGTCCATTGAATGTGAAAAGAGTTGCAGAATCTTGAAAAGCAGCTTGAGGGAGCTCTTCTCCAGGCTAGGCAAAGG	640
641	<mark>K T Q L M I E Q M E E L R R K E R Q L G D L N K Q L K</mark> AAGACTCAGCTAATGATTGAACAAATGGAAGAGGCTCCGAAGAAAGGAGCGGCAGCTTGGGGACTTGAATAAGCAGCTTAA	720
721	<mark>I K V</mark> S L E M S S L E A A E G Q G L I R G L P W L W GATTAAGGTTTCCCTGGAAATGTCATCGCTTGAAGCAGCTGAGGGACAAGGCTTAATAAGAGGGCCTCCCTTGGCTTTGGA	800
801	S S S V P S G S S M F P M H T S H P S A A M D C D P E GTTCGAGTGTACCAAGTGGAAGCAGCATGTT <u>TCCTATGCACACTTCTCACC</u> CCAGTGCCGCCATGGATTGCGATCCTGAG	880
881	PVLQIGYHQYAPAEGPSAPRSMAIESN CCTGTCCTACAGATAGGGTATCATCAGTATGCTCCGGCAGAAGGTCCTTCTGCCCCGAGGAGCATGGCCATTGAGAGTAA	960
961 1041 1121 1201	I I Q G W A L * Y L F A N L L Y V * C K K P I Q * F CATCATCCAGGGGTGGGCTCTTTGATATCTCTTTGCAAATCTTTTATACGTTTGATGCAAAAAACCAATTCAATGATTG TGCTTTTGGATTTGAACATGTTAATCCCATATGAAACTAAGTACTTTATAACTAGCTGCTTAATTTTGTATTTACAGAAC TTGAAACTTACGCATTTTGATCTTTATTGCAAGGATTATAATCGCATGATGTAGACTGTTTGAATCTTAAAAAAAA	1040 1120 1200

22) CaSHP (Coffea arabica SHATTERPROOF) - Incompleta

Reads: CA00-XX-FB2-018-E01-JF.F CA00-XX-FB2-022-F10-JE.F (clone) CA00-XX-FB4-041-D07-AG.F (clone)

```
>Proteína
```

YEYANNSVRGTIERYKKACADSSNPGSVSEANAQFYQQEASKLRKHIREIQNSNRHILGDGVDGLNFKELKNLEGKVEKAIGRIRTRKNELLFA EIELMHKREIELQNANTYLRAKIAENERAQQHMNLMPGSEYQPLASQPYDVRNFLPVNLLEPDQHYSRQDQTALQLV-

		Y	Е	Y		A	N	N S	s t	V	R	G	Т	Ι	Е	R	Y	Κ	K	А	С	А	D	5	5 5	1 3	J	Р	G	
1	ΤT	TAC	GA	ATA	ΤG	СТА	ATA	ACA	GTG	ГСА	GGG	GGA	CCA	ATTO	GAGA	GGT	ACA	AG	AAA	GCC	TGI	GC	ΓGA	TTC	CATO	CAA	4TC	CA	GGA	80
81	S TC	V TGT	TT(S CTG	E AA(A GCT	N AAC	A GCC(Q CAGI	F TTC	Y TAT	Q CAG	Q CAA	E AGA <i>l</i>	A AGCI	<mark>S</mark> AGC	<mark>K</mark> AAG	L GCT(R CAG	<mark>K</mark> GAA	H AC	I ATA:	I Faa	<mark>r</mark> Gac	E GAG <i>i</i>	I ATAC	Q CAG	<mark>N</mark> GAA	<mark>S</mark> ITC	160
161	AA	<mark>n</mark> Aca	<mark>R</mark> .GA(H CAC	I AT	L CCT	<mark>G</mark> GGG	D TGA	<mark>G</mark> CGG(V GGT	AGA	<mark>) G</mark> ATGG	L ATI	AA?	<mark>V F</mark> ACTI	CAA	E AGA	AC'	L TGA	<mark>k</mark> Aga	<mark>n</mark> Atc	L CTG(<mark>e</mark> Gaa	G GG1	K AA	V AGT(E GGA	GA2	<mark>k</mark> Aag	240
	A	Ι	G	R		I	R	ΤI	RI	K	N	E	L	L	F	A	E	Ι	Е	L	М	H	K	F	ξĒ	6 3	[E	L	

241	CCATTGGCAGAATCCGTACCAGGAAGAACGAGTTATTGTTTGCTGAAATCGAGCTTATGCATAAGAGGGAAATTGAGCTG	320
321	<mark>Q N A N T Y L R A K I A E</mark> N E R A Q Q H M N L M P G S CAAAATGCCAATACGTACCTGAGAGGCAAAGATTGCAGAAAAGGAAGAGCACAGCAGCACATGAACTTAATGCCAGGGTC	400
401	E Y Q P L A S Q P Y D V R N F L P V N L L E P D Q H TGAATACCAGCCTCTTGCTTCACAGCCTTATGATGTTAGGAACTTCCTCCCAGTAAACCTCCTGGAACCTGATCAACATT	480
	Y S R Q D Q T A L Q L V * D L P A E G K C F P L D R P	
481	ACTCTCGCCAGGACCAAACTGCTCTTCAACTCGTCTAAGATTTGCCTGCTGAAGGGAAGTGTTTTCCATTGGATAGGCCT	560
561	CTATTTTACTAGCACTATGGCAAGAGCTCTTGTCAATTTATGAGAACAAGATATATGTAGCTATAACTAAC	640
641	GAGTATGTTGTGTATAAGAATGATCTCGTTTTACTTGTGGACAACATATCTATTTTCACCCACGCGTCCGAAAAAAAA	720
721	ААААААААААААААААААААА 742	

Item 3 – Série de gemas e botões florais com mucilagem



Obs.: Fotos tiradas em Lupa. A mucilagem dá a coloração ocre-amarelada ao topo das gemas.



Itens 4 - Gráficos níveis de expressão dos ortólogos ABC durante desenvolvimento floral



Obs.: Gráfico 1 mostra a inversão dos níveis de expressão dos genes *13-CaAP1* (A) e *5-CaAP3/14-CaTM6* (B) envolvidos na formação das sépalas e pétalas e, o Gráfico 2 mostra o aumento da expressão dos genes *10-CaPI* e *12-CaAG* envolvidos na formação dos estames e carpelos, ao longo do desenvolvimento floral. As intensidades de expressão foram quantificadas pela ferramenta Density Tool do Gel doc nas RT-PCR específicas dos estádios de desenvolvimento floral (Figura 8A). Para cada um desses genes e os valores normalizados em relação a intensidade de seus respectivos controles negativos.