



GABRIEL DE CAMPOS RUME

**ANÁLISE GLOBAL DOS GENES MADS-BOX ENCONTRADOS
NO GENOMA DE *Coffea arabica* L. E SUA ORIGEM A PARTIR
DOS PARENTAIS DIPLÓIDES (*C. canephora* E *C. eugenioides*)**

**LAVRAS – MG
2021**

GABRIEL DE CAMPOS RUME

**ANÁLISE GLOBAL DOS GENES MADS-BOX ENCONTRADOS NO GENOMA DE *Coffea arabica* L. E SUA ORIGEM A PARTIR DOS PARENTAIS DIPLÓIDES
(*C. canephora* E *C. eugenioides*)**

Monografia de conclusão de curso apresentada à Universidade Federal de Lavras, como parte das exigências do Curso de Ciências Biológicas, para a obtenção do título de Bacharel.

Prof. Dr. Antônio Chalfun Júnior
Orientador

Dr. Raphael Ricon de Oliveira
Coorientador

**LAVRAS – MG
2021**

**Ficha catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca
Universitária da UFLA, com dados informados pelo(a) próprio(a) autor(a).**

Rume, Gabriel de Campos.

Análise global dos genes MADS-box encontrados no genoma de Coffea arabica L. e sua origem a partir dos parentais diplóides (C. canephora e C. eugenioides) / Gabriel de Campos Rume. - 2021.
22 p. : il.

Orientador(a): Antônio Chalfun Júnior.

Coorientador(a): Raphael Ricon de Oliveira.

TCC (graduação) - Universidade Federal de Lavras, 2021.

Bibliografia.

1. Coffea arabica. 2. MADS-box. 3. Desenvolvimento Vegetal.
I. Júnior, Antônio Chalfun. II. Oliveira, Raphael Ricon de. III.
Título.

GABRIEL DE CAMPOS RUME

**ANÁLISE GLOBAL DOS GENES MADS-BOX ENCONTRADOS NO GENOMA DE *Coffea arabica* L. E SUA ORIGEM A PARTIR DOS PARENTAIS DIPLÓIDES
(*C. canephora* E *C. eugenioides*)**

**GENOME-WIDE ANALYSIS OF THE MADS-BOX GENE FAMILY IN THE
ALLOTETRAPLOID *Coffea arabica* L. AND ITS ORIGIN FROM THE DIPLOID
PARENTAL SPECIES (*C. canephora* AND *C. eugenioides*).**

Monografia de conclusão de curso apresentada à
Universidade Federal de Lavras, como parte das
exigências do Curso de Ciências Biológicas, para a
obtenção do título de Bacharel.

APROVADO em 10 de junho de 2021.
Prof. Dr. Antônio Chalfun Júnior UFLA
Dr. Carlos Henrique Cardon UFLA
MS. Thales Henrique Cherubino UFLA
Dr. Raphael Ricon de Oliveira UFLA

Prof. Dr. Antônio Chalfun Júnior
Orientador

Dr. Raphael Ricon de Oliveira
Coorientador

**LAVRAS – MG
2021**

*Ao meu pai Francisco e à minha mãe Sandra,
por toda a amizade, apoio, amor e exemplo.
Sem vocês nada disso seria possível.*

AGRADECIMENTOS

À maravilhosa Universidade Federal de Lavras, ao Departamento de Biologia e a meus professores ao longo desses anos, por todas as oportunidades e contribuições ao meu desenvolvimento pessoal e acadêmico.

Ao meu pai Francisco, minha mãe Sandra, minha irmã Júlia e à todos meus familiares, por sempre proporcionarem tanto apoio, amor e incentivo.

À todas as amigas que tive o privilégio de construir desde meu primeiro dia em Lavras, em especial à República MethioLate e aos amigos da turma do 16/2 de ciências biológicas, por todos os momentos de alegria, apoio e crescimento ao longo dos anos.

Às pessoas que foram essenciais para a realização desse trabalho: meu orientador e Professor Dr. Antônio Chalfun, meu coorientador Dr. Raphael Ricon e o doutorando Thales Cherubino. Assim como aos amigos e colegas do laboratório de Fisiologia Molecular de Plantas da UFLA, por todo o exemplo, experiência, e conhecimento fornecidos em meu período de IC.

Muito obrigado!

RESUMO

A família MADS-box de fatores de transcrição possui importantes funções em diversos aspectos dos desenvolvimentos vegetativo e reprodutivo das plantas. Suas funções foram estudadas em plantas modelo e estão em grande parte relacionadas à determinação da identidade meristemática e ao desenvolvimento de flores. Uma análise global da família MADS-box ainda não havia sido realizada em *Coffea arabica* L., uma importante commodity agrícola e recente espécie poliplóide, única no gênero. Portanto, um estudo mais completo dos genes MADS-box auxiliará na compreensão de aspectos evolucionários dessa família de fatores de transcrição, sendo útil também para programas de melhoramento. Aqui, nós identificamos um total de 102 genes MADS-box, dos quais 81 são completamente novos e 21 são sequências já publicadas, validando nosso método de busca. Através de análises comparativas usando proteínas de *Arabidopsis thaliana*, árvores filogenéticas foram geradas e classificamos as proteínas identificadas em seus respectivos subgrupos, com o grupo MIKC sendo adicionalmente categorizado em 16 subfamílias. Um mapa cromossômico contendo os genes foi elaborado, onde pudemos observar uma distribuição irregular e baixa variação posicional de genes entre cromossomos homeólogos de cada espécie parental (*C. canephora* e *C. eugenioides*). Foi também realizada uma análise estrutural com o intuito de investigar a ocorrência de variações e confirmar as classificações atribuídas. A partir da análise foi verificado que o número de íntrons nos genes MADS-box varia de 0 a 11, e pudemos confirmar a relação tipo-estrutura na maioria dos genes identificados, com algumas exceções. Além disso, é possível também observar uma alta similaridade estrutural entre genes proximamente relacionados, ocorrendo variação apenas no comprimento de íntrons. A fim de fornecer informações sobre as funções dos genes MADS-box, nas próximas etapas determinaremos sua expressão em diferentes tecidos do cafeeiro por meio de análises *in silico* de RNA-Seq. Portanto, nosso trabalho contribui na elucidação da história evolutiva da família MADS-box em Rubiaceae e no direcionamento de estudos de desenvolvimento no gênero *Coffea*.

Palavras-chave: *Coffea arabica* L.; MADS-box; Fator de Transcrição; Florescimento; Desenvolvimento Vegetal; Análises Filogenéticas.

SUMÁRIO

1.	INTRODUÇÃO	8
2.	MATERIAL E MÉTODOS	10
2.1.	Identificação e Anotação dos Genes	10
2.2.	Análises Filogenéticas	10
2.3.	Mapeamento dos genes MADS-box nos cromossomos e análise estrutural	11
3.	RESULTADOS.....	11
3.1.	Identificação e classificação dos genes MADS-box de <i>Coffea arabica</i>	11
3.2.	Mapa da posição nos cromossomos e Estrutura dos genes MADS-box do cafeeiro	14
4.	Considerações Finais.....	16
	APÊNDICE A	17
	REFERÊNCIAS	19

Artigo a ser submetido e formatado de acordo com as normas da revista Genes.

Análise global dos genes MADS-box encontrados no genoma de *Coffea arabica* L. e sua origem a partir dos parentais diplóides (*C. canephora* e *C. eugenioides*)

Gabriel de C. Rume ¹, Thales H. C. Ribeiro ¹, Raphael R. de Oliveira ¹ e Antônio Chalfun-Junior ^{1,*}

¹ Departamento de Biologia, Universidade Federal de Lavras, Lavras – MG, Brasil

* Autor Correspondente: chalfunjunior@ufla.

Abstract: The MADS-box family of transcription factors has important functions in various aspects of the vegetative and reproductive cycles of plant development, their functions are studied in model plants and are mostly related to the determination of meristem identity and flower development. A global analysis of the MADS-box family was not yet performed in *Coffea arabica* L., an important agricultural commodity and a recently formed polyploid species, unique in the genre. Thus, a more complete study of MADS-box genes in crops would help understand evolutionary aspects of this family of transcription factors, while being useful to breeding programs. Here, we identified a total of 102 MADS-box genes, from which 81 are totally new and 21 are already published sequences, validating our search method. Through comparative analyses using *Arabidopsis thaliana* proteins we generated phylogenetic trees and classified the identified proteins from coffee in their respective subgroups, with the MIKC^c group being further categorized into 16 subfamilies. A chromosomal map containing the genes was elaborated, where we could observe a distribution irregularity and low positional variation of genes between homeolog chromosomes of each parental species (*C. canephora* and *C. eugenioides*). Moreover, a structural analysis revealed that the number of introns in coffee MADS-box genes ranges from 0 to 11, and confirmed the gene type-structure relationship in most of the identified genes. Moreover, we also observed that closely related genes have very high structural similarity, with variations only in intron length. In order to provide insights about MADS-box genes functions, in the next steps we will determine their expression in different coffee tissues through *in silico* RNA-Seq analysis. Thus, our work contribute to elucidate the MADS-box family evolutionary history in Rubiaceae and to direct development studies in the *Coffea* genre.

Keywords: *Coffea arabica*; MADS-box; Transcriptional Factor; Flowering; Plant Development

1. Introdução

Pertencentes à família Rubiaceae e ao gênero *Coffea*, com 103 espécies descritas [1], *Coffea arabica* L. e *Coffea canephora* são as duas espécies cultivadas das quais as sementes são utilizadas na produção do café, considerada importante commodity agrícola [2]. *Coffea arabica* é a única espécie poliplóide do gênero, sendo caracterizada como uma alotetraplóide ($2n=4x=44$) originada da hibridação de gametas não reduzidos de *Coffea canephora* e *Coffea eugenioides* [3,4]. O cafeeiro é uma planta bianual, ou seja, seu desenvolvimento ocorre em ciclos de dois anos, iniciados após o término da fase juvenil [5]. O primeiro ano é caracterizado pelo desenvolvimento de ramos e folhas juntamente à indução e desenvolvimento reprodutivo, e o segundo ano inicia-se com a antese, desenvolvimento de frutos e senescência de ramos [5]. Assim como em outras plantas, seu desenvolvimento depende de uma complexa rede de regulação genética que integra diversos estímulos endógenos e exógenos, como estresse hídrico, temperatura e níveis de açúcar, com receptores e integradores de sinais através da ativação coordenada de Fatores de Transcrição (FT), proteínas nucleares que impulsionam ou inibem a transcrição de genes [6].

A família MADS-box de FT está relacionada a diversos processos nos organismos, como respostas fisiológicas em fungos, crescimento de músculos em animais e no desenvolvimento de flores em plantas [7]. Entretanto, o número de genes MADS-box presentes nos genomas de diferentes linhagens varia consideravelmente, sendo menor em protistas e animais e atingindo as centenas em algumas plantas fanerógamas [7]. MADS é um acrônimo derivado dos nomes dos quatro genes que fundaram essa família: *MINICHROMOSOME MAINTENANCE (MCM1)* de *Saccharomyces cerevisiae*, *AGAMOUS (AG)* de *Arabidopsis thaliana*, *DEFICIENS (DEF)* de *Antirrhinum majus* e *SERUM RESPONSE FACTOR (SRF)* de *Homo sapiens* [8]. A presença de tal família em organismos tão distantes evidencia sua ancestralidade, enquanto sua conservação representa a sua importância na regulação transcricional dos processos de desenvolvimento [7]. Contudo, tal família não teve todos os seus membros descritos no gênero *Coffea* (Rubiaceae), sendo encontrados apenas poucos trabalhos, relacionados principalmente ao florescimento [9–11].

Particularmente nas plantas, os FT MADS-box possuíram papel fundamental no processo evolutivo, atuando no controle de diversos mecanismos essenciais para que os organismos completem seu ciclo de vida [12]. Diversos estudos já atribuíram à família funções em diversos aspectos do desenvolvimento em plantas, como por exemplo, seu papel na determinação da identidade em células meristemáticas e do tempo de florescimento, na fertilidade de grãos de pólen e no desenvolvimento dos ovários e frutos [13]. Além disso, também já foram encontrados padrões de expressão em folhas, células-guarda e tricomas, porém sua função exata nesses tecidos ainda é desconhecida [9].

Os membros da família MADS-box são classificados em dois tipos principais, originados por uma duplicação ocorrida antes da divergência entre plantas e animais, sendo distinguidos principalmente pelos domínios conservados presentes nas proteínas [14]. Outras características distintivas entre subgrupos MADS-box são o número de éxons e suas diferentes taxas de nascimento e morte [7]. Até o momento foram caracterizados cinco subgrupos de genes MADS-box, sendo três pertencentes ao tipo I ($M\alpha$, $M\beta$ e $M\gamma$), enquanto os outros dois estão agrupados no tipo II (MIKC^c e MIKC*) [15].

Genes MADS do tipo I (*SRF-like*) são caracterizados por possuírem apenas o domínio MADS que está situado na região amino-terminal da proteína e é descrito como responsável pela ligação com o DNA em áreas específicas localizadas nos sítios promotores dos genes alvo [14]. Tais genes podem ser subdivididos em três grupos: $M\alpha$, $M\beta$ e $M\gamma$, com base em sua filogenia e na presença de motivos conservados de aminoácidos na região carboxi-terminal [7]. Estudos funcionais sobre este grupo de genes são escassos, mas as funções descritas até hoje estão relacionadas principalmente ao desenvolvimento do gametófito feminino, do saco embrionário e de sementes [16]. Por exemplo, em *Arabidopsis thaliana*, foi comprovado que a ausência do gene *AGAMOUS-LIKE 23 (AGL23)*, subgrupo $M\alpha$, resultava no impedimento do desenvolvimento do gametófito feminino, e na persistência do megásporo durante as fases subsequentes do desenvolvimento do óvulo [7].

Já as proteínas classificadas no tipo II (*MEF2-like*), chamadas também de proteínas MIKC, são identificadas pela presença, além do domínio MADS, de um domínio I (*intervening*) responsável pela especificidade da ligação com o DNA, de um domínio K (*keratin-like*) que facilita a dimerização das proteínas, e um domínio C (*carboxi-terminal*) menos conservado que é descrito como necessário em algumas interações para a ativação da transcrição dos genes alvo [17]. As proteínas MIKC são também subdivididas em dois grupos: MIKC^c e MIKC*, distinguidos por diferenças na estrutura do domínio K e no comprimento e número de éxons codificando o domínio I [7]. O conhecimento que possuímos sobre os genes MIKC é significativamente maior quando comparado ao tipo I, e suas funções estão relacionadas principalmente ao desenvolvimento do esporófito, à especificação da identidade floral e à regulação do tempo de florescimento em resposta à estresses [16].

O papel dos genes MADS-box do tipo II foi bastante explorado no desenvolvimento reprodutivo, sendo descrito um modelo conhecido como ABC que descreve a determinação da identidade floral, importante para a compreensão de aspectos evolutivos das plantas [18,19]. O modelo ABC descreve que a identidade floral é controlada através da combinação das funções de três classes de genes [6].

Correspondentes à classe A: *APETALA 1 (AP1)* e *FRUITFULL (FUL)*. *APETALA 3 (AP3)* e *PISTILLATA (PI)* pertencem à classe B. *AGAMOUS (AG)* representa a classe C. A expressão desses genes em cada camada do órgão floral determina o verticilo, ou seja, A especifica sépalas, A e B juntos direcionam a identidade das pétalas, B e C determinam a identidade de estames, e C controla a identidade dos carpelos. Foi também descrito que os genes *SEEDSTICK (STK)* e *SHATTERPROOF1 (SHP1)* são necessários para a identidade dos óvulos, e incluídos no modelo como genes da função D [20,21], porém, a conservação de tal função ainda não foi demonstrada em outros grupos vegetais. Posteriormente, foi descrito que as proteínas MADS-box atuam em complexos quaternários (*Floral quartets*) sendo necessária a presença dos genes *SEPALLATA (SEP)* para que ocorra essa ligação [22]. Os genes *SEP* também pertencem ao grupo MADS-box e, por atuarem de maneira conservada, foram denominados genes de classe E, constituindo o modelo ABCE expandido [18].

Estudos anteriores já identificaram e analisaram alguns dos genes MADS-box presentes em *C. arabica*, buscando a compreensão de seu envolvimento no desenvolvimento reprodutivo da planta [9–11]. Porém, devido à grande importância dessa família de fatores de transcrição e à disponibilização recente do genoma completo de *C. arabica* (NCBI *Coffea arabica* Annotation Release 100), realizamos uma análise global *in silico* para a identificação e caracterização de todos os putativos genes MADS-box encontrados no genoma do cafeeiro. Como resultado, 102 genes foram identificados e agrupados nas diferentes subclasses da família MADS-box, suas posições foram determinadas no genoma e suas estruturas demonstradas. Sendo assim, o presente estudo busca auxiliar na compreensão dos mecanismos moleculares que coordenam processos dos desenvolvimentos vegetativo e reprodutivo do cafeeiro, através da identificação e classificação de fatores importantes, tais como os representantes da família MADS-box. Sendo *Coffea arabica* uma espécie relativamente recente, formada da hibridização entre *C. eugenioides* e *C. canephora*, e também uma espécie de grande relevância econômica, discutimos aspectos evolutivos, bem como possibilidades de futuros estudos para o melhoramento dessa cultura.

2. Material e Métodos

2.1. Identificação e Anotação dos Genes

Sequências proteicas preditas de *Coffea arabica* foram obtidas da anotação de seu genoma e proteoma disponíveis no site do National Center for Biotechnology Information (NCBI, disponível em https://www.ncbi.nlm.nih.gov/assembly/GCF_003713225.1/#/def) e utilizadas na criação de um banco de dados local. Foram feitos downloads, do banco de dados TAIR (<https://www.arabidopsis.org/>) [23], de 107 proteínas MADS-box já descritas de *Arabidopsis thaliana*. Também foram utilizados representantes da família de fatores de transcrição em espécies de tomate (gênero *Solanum*) e no arroz (*Oryza sativa*), obtidas das páginas Sol Genomics (<https://solgenomics.net/>) e Rice Genome Annotation Project (<http://rice.plantbiology.msu.edu/>), respectivamente.

Para a identificação das sequências de interesse, proteínas MADS-box anotadas de *Arabidopsis* foram utilizadas em alinhamentos contra o proteoma de *C. arabica* com uso da ferramenta BLASTp v2.10.1 [24], tendo como parâmetro um e-value mínimo de 10^{-3} . Exaustivas filtragens foram realizadas sobre as proteínas identificadas, em que removemos das análises subsequentes todas as proteínas identificadas no cafeeiro que possuíam comprimento 70% maior ou 70% menor do tamanho médio das proteínas de *Arabidopsis* utilizadas como referência, e manualmente removemos suas diferentes isoformas, de maneira a restar apenas uma representante (unigenes) por gene MADS-box. Além disso, as sequências candidatas restantes foram analisadas através de buscas por seus domínios conservados no banco de dados PFAM (<http://pfam.xfam.org/>) [25] e pela ferramenta SMART (<http://smart.embl-heidelberg.de/>) [26], sendo a presença do domínio MADS confirmada em todas elas.

2.2. Análises Filogenéticas

Para a classificação dos genes MADS-box de *Coffea arabica*, realizamos um alinhamento global das proteínas candidatas restantes com as sequências proteicas já

curadas de *Arabidopsis*, tomate (gênero *Solanum*) e arroz (*Oryza sativa*), com uso do programa MAFFT v7.475 [27] através do algoritmo Guidance v2.02 [28]. Árvores filogenéticas foram geradas com algoritmos contidos no pacote PHYLIP [29] sendo utilizado o seguinte procedimento para cada uma delas: Um *dataset* de 1000 randomizações (*bootstraps*) foi gerado com o algoritmo *seqboot*, as matrizes de similaridade entre as diferentes sequências foram calculadas com o algoritmo *protdist* e a construção das árvores realizada através do método *neighbor-joining* com o algoritmo *neighbor*. Por fim, a elucidação das árvores foi feita com o algoritmo *consense* e o programa FigTree v1.4.4 (<http://tree.bio.ed.ac.uk/software/figtree/>) utilizado para a visualização das árvores e obtenção das figuras.

2.3. Mapeamento dos genes MADS-box nos cromossomos e análise estrutural

Dados de localização e estrutura correspondentes a cada gene MADS-box identificado foram obtidos a partir do arquivo GFF disponibilizado com a anotação do genoma de *C. arabica*. O mapa físico foi desenhado com a ferramenta MapGene2Chrom v2.0 (<http://mg2c.iask.in/mg2c%5Fv2.0/>) [30]. Não foi possível determinar as posições de 15 genes pois correspondiam a sequências genômicas ainda não organizadas do genoma da planta.

As figuras que ilustram as estruturas dos genes foram geradas pela ferramenta GSDS v2.0 (<http://gsds.gao-lab.org/>) [31].

3. Resultados e Discussão

3.1. Identificação e classificação dos genes MADS-box de *Coffea arabica*

Como resultado de buscas no proteoma através da ferramenta BLASTp e após a realização de filtragens, remoção de isoformas e confirmação da presença do domínio MADS, um total de 102 putativos genes MADS-box foram identificados em *Coffea arabica* e nomeados de CaMADS1 a CaMADS102, apresentados de forma resumida na Tabela 1. Das 102 sequências encontradas neste trabalho, 81 são novas e outras 21 são similares à sequências previamente publicadas [10], conforme indicado na Tabela 1 do Apêndice (Tabela A1).

Tabela 1. Número de sequências representantes de cada subgrupo da família MADS-box, identificadas em *C. arabica*.

	Subgrupo	Nº
Tipo I	M α	31
	M β	0
	M γ	11
Tipo II	MIKC ^c	55
	MIKC*	5
Total		102

A partir da verificação da presença do motivo conservado K-box e utilizando como referência as proteínas MADS de *Arabidopsis* reportadas por Parenicová et al. [12], foi possível a montagem de árvores filogenéticas e classificação dos genes MADS-box do cafeeiro nos diferentes subgrupos descritos. Do total de 102 genes MADS-box, 42 foram classificados como tipo I, sendo 31 representantes do subgrupo M α e 11 do subgrupo M γ , e 60 foram classificados como tipo II, sendo 55 pertencentes ao subgrupo MIKC^c e 5 ao MIKC* (Figura 1). Não identificamos nenhum gene M β de *C. arabica* em nossas

análises, porém essa peculiaridade também já foi reportada em estudos do mesmo tipo realizados em outras espécies de plantas [32].

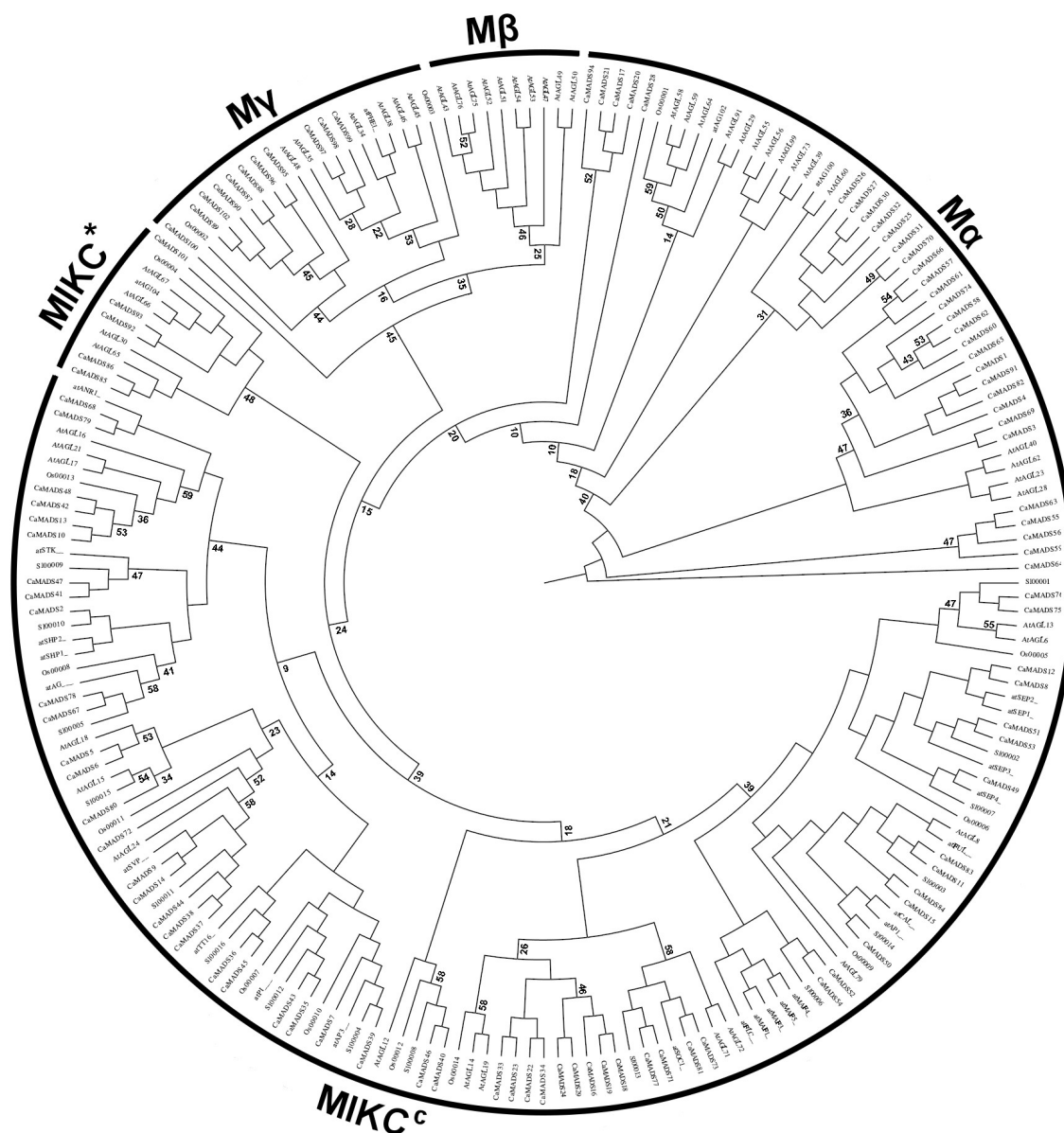


Figura 1. Árvore filogenética gerada através do método *neighbor-joining* contendo proteínas de *Arabidopsis* (81), Arroz (14), Tomate (16), e todas as seqüências identificadas em *C. arabica*. Ao redor da árvore estão indicados os respectivos subgrupos a que pertencem. Nós com valores de *bootstrap* menores que 60% e que, portanto, não apresentam suporte estatístico, foram indicados na figura.

No intuito de classificar mais especificamente as proteínas tipo II do cafeeiro, uma segunda árvore filogenética foi gerada contendo apenas seqüências **Mikc^c** de *C. arabica* e *Arabidopsis* (Figura 2). De acordo com o alto valor de *bootstrap* com a qual as proteínas se

agruparam na árvore, identificamos os homólogos e prováveis ortólogos dentro dos subgrupos já descritos. Essa análise permite, juntamente com futuras análises de expressão, realizar inferências baseadas em homologia com genes MADS-box já caracterizados e cujas funções estão relacionadas desde o desenvolvimento de raízes até o controle do tempo de florescimento e na arquitetura das flores [7,33].

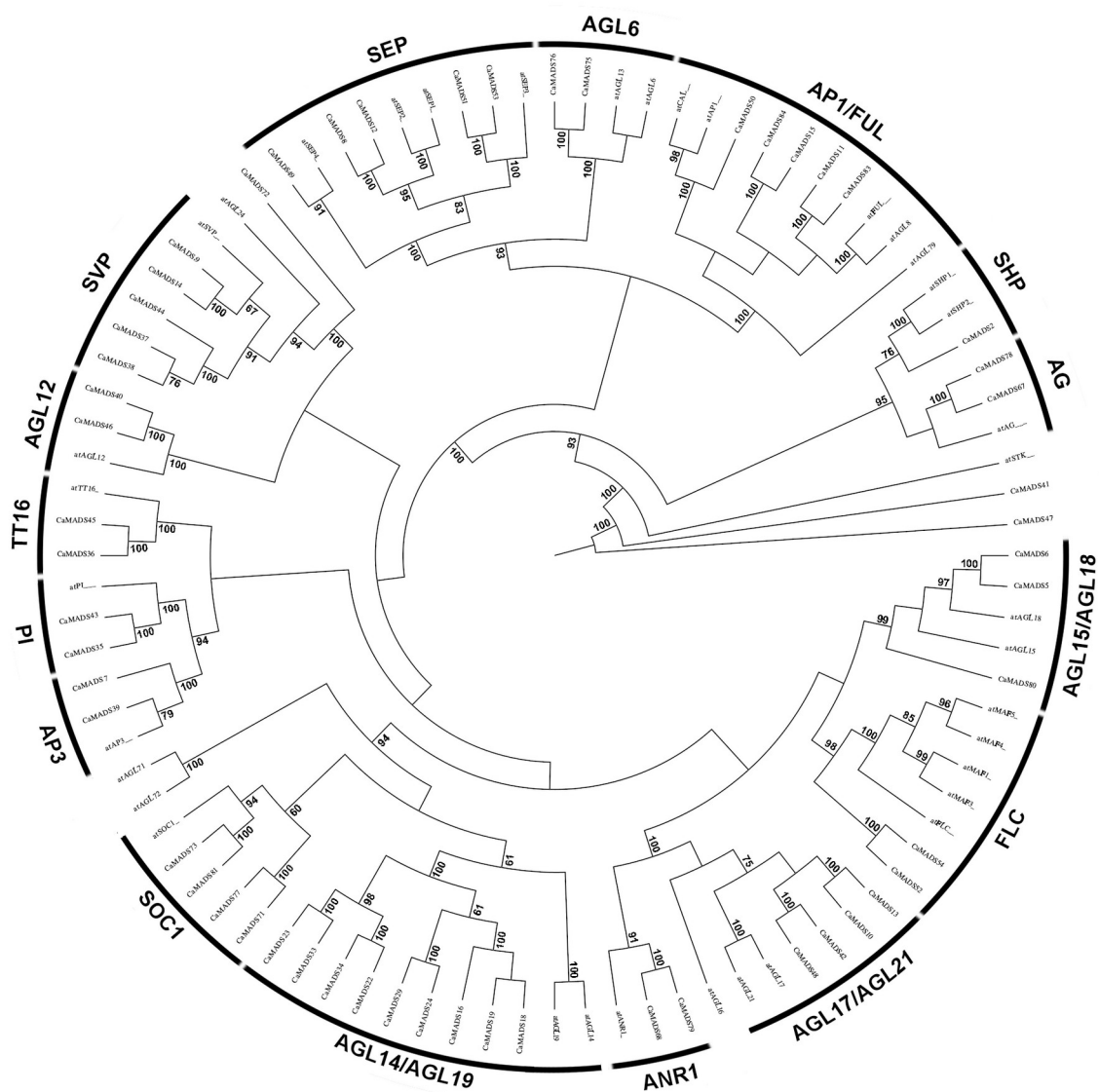


Figura 2. Árvore filogenética gerada através do método *neighbor-joining* contendo apenas proteínas MIKC de *Arabidopsis* e *C. arabica*. Ao redor da árvore estão indicados os respectivos subgrupos a que pertencem. Nós com valores de *bootstrap* maiores que 60% e que, portanto, apresentam suporte estatístico, foram indicados na figura.

Devido à natureza poliplóide de *C. arabica*, foram identificados possíveis múltiplos ortólogos para diversos genes MIKC de *Arabidopsis*. Por exemplo, dois homólogos para os genes *AGAMOUS* (*AG*), *SEPPALATA3* (*SEP3*), *AGAMOUS-LIKE12* (*AGL12*), *TRANSPARENT TESTA16* (*TT16*), *PISTILLATA* (*PI*), *APETALA 3* (*AP3*), *ARABIDOPSIS NITRATE REGULATED 1* (*ANR1*) e *AGL18* puderam ser identificados de maneira inequívoca no genoma do cafeeiro. Curiosamente, os genes *FRUITFULL* (*FUL*), *SHORT VEGETATIVE PHASE* (*SVP*), *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1*

(*SOC1*), *AGL14* e *AGL19* aparentam possuir ainda mais representantes, chegando a quatro.

Outros genes importantes, como *FLOWERING LOCUS C (FLC)*, *AP1*, e *AGL6* foram classificados dentro dos subgrupos descritos, servindo como um controle positivo para as análises. Entretanto, devido à distância evolutiva entre *C. arabica* e *Arabidopsis*, tais sequências foram agrupadas sem suporte estatístico (valores de *bootstrap* menores que 60%) em relação aos seus prováveis ortólogos. Outras árvores considerando espécies mais próximas do cafeeiro poderiam resolver a ortologia, porém não foram realizadas uma vez que o objetivo do trabalho foi o de classificar de maneira geral todas as sequências encontradas no genoma de *C. arabica*.

3.2. Mapa da posição nos cromossomos e Estrutura dos genes MADS-box do cafeeiro

A partir de dados contidos no arquivo GFF disponibilizado com a anotação do genoma (NCBI *Coffea arabica* Annotation Release 100), foram posicionados todos os genes MADS-box nos cromossomos de cafeeiro, exceto para 15 genes ainda não organizados no genoma da planta. Assim, foi possível a elaboração de um mapa contendo suas posições nos cromossomos do cafeeiro (Figura 3). Através da visualização do mapa podemos destacar a irregularidade na distribuição dos genes MADS-box no genoma de *C. arabica*. Por exemplo, não é observada a presença de nenhum gene no 3º par cromossômico bem como em um dos cromossomos do 5º par, enquanto os pares 2, 7, 9 e 10 contém acima de 50% de todos os genes identificados.

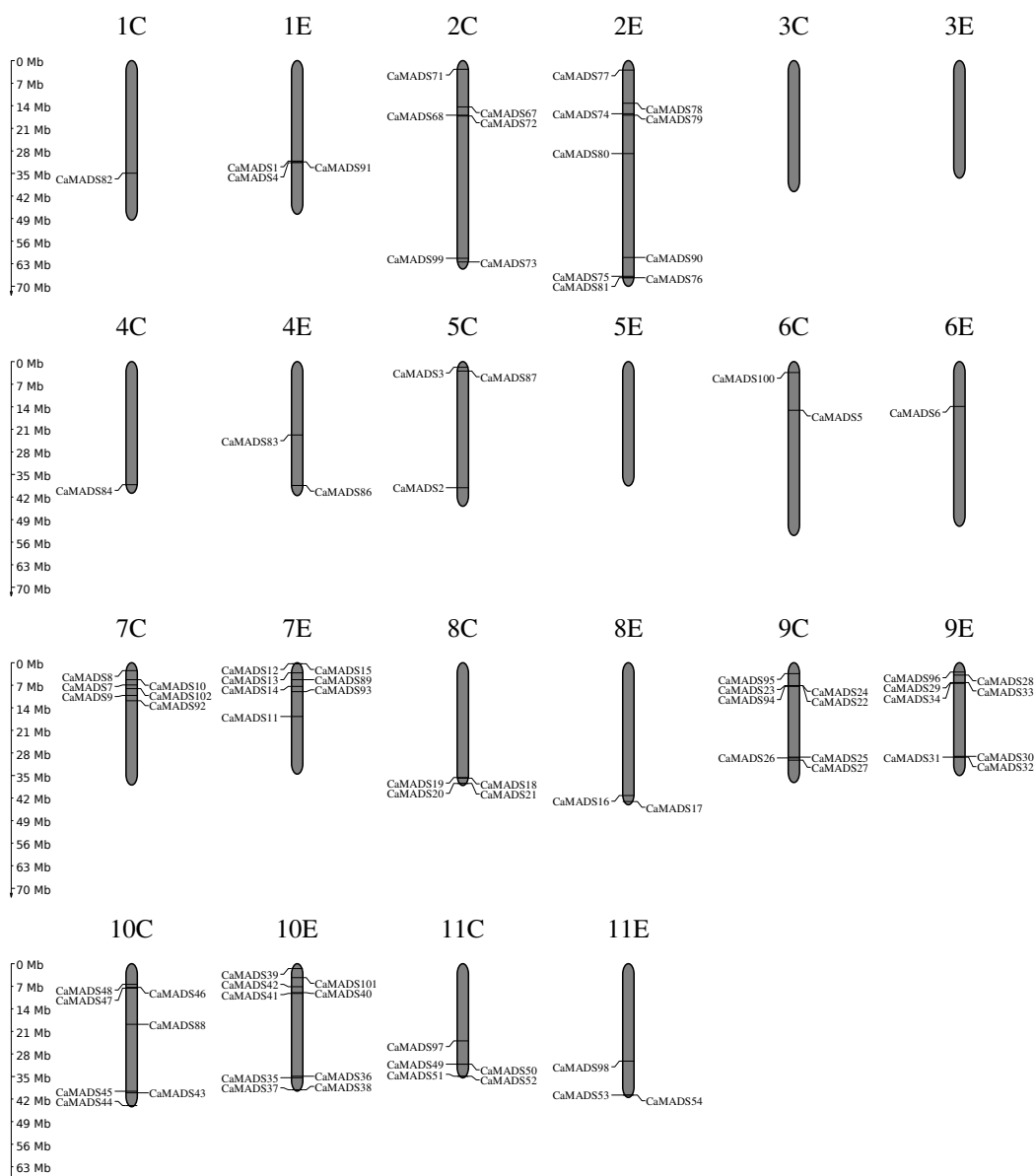


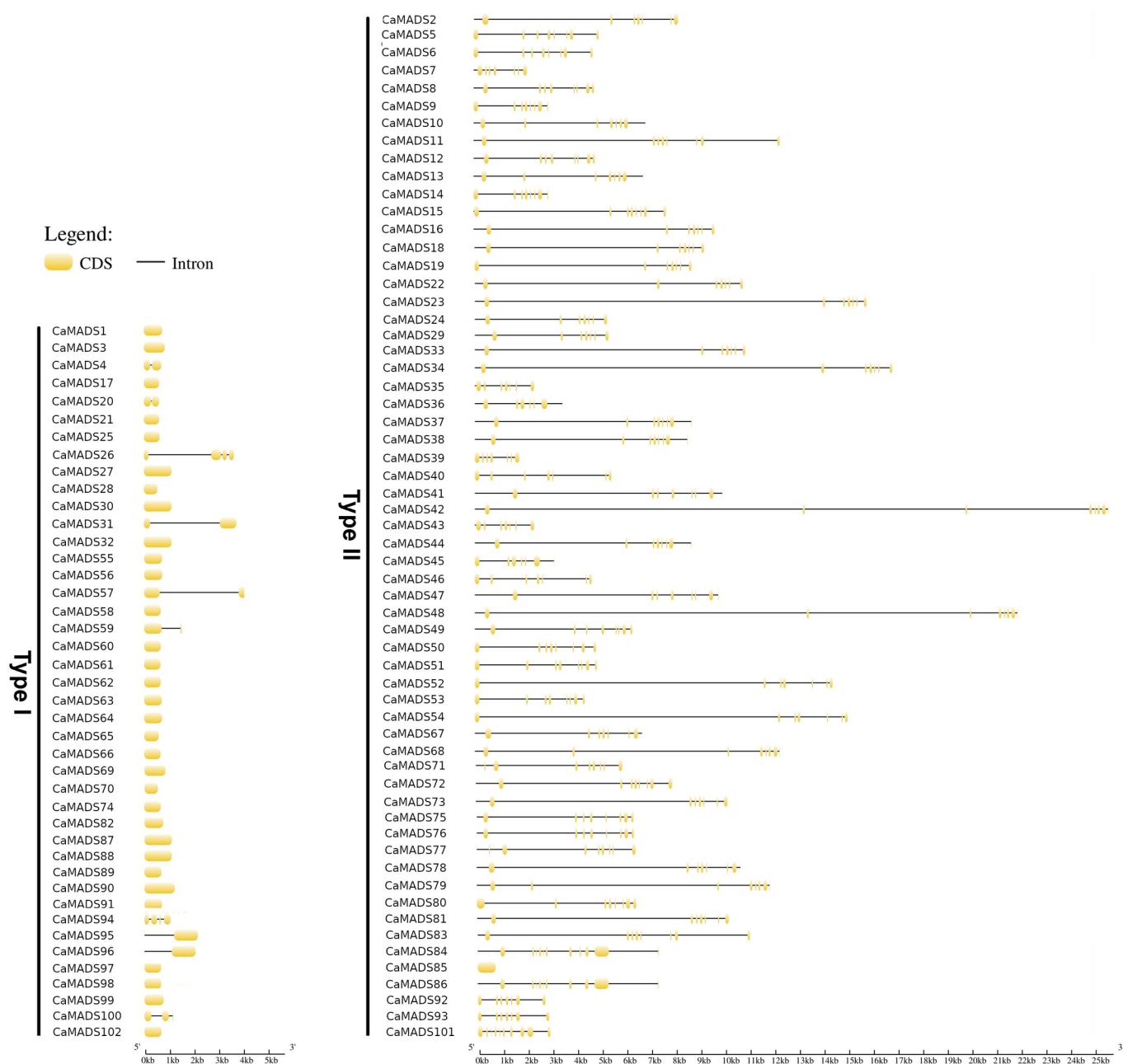
Figura 3. Distribuição dos putativos genes MADS-box nos cromossomos de *C. arabica*. Os cromossomos estão identificados pelo seu número (1-11) seguido por C ou E, representando o parental de que provém, *C. canephora* ou *C. eugenioides*, respectivamente. A posição dos genes e o tamanho dos cromossomos podem ser estimados através da escala à esquerda da figura, em Megabases (Mb).

C. arabica é um anfidiplóide relativamente recente originado do cruzamento interespecífico entre *C. canephora* e *C. eugenioides* [3]. *C. arabica* possui um conjunto completo de cromossomos de cada espécie parental e é, portanto, caracterizado como um alotetraplóide [3]. Eventos de deleção de sequências homeólogas, pares de genes originados por especiação e reunidos novamente no mesmo genoma por aloploidização, é algo comum na história evolutiva das plantas [34]. Entretanto, devido à ausência de tempo suficiente (no contexto evolutivo) e ao fato de *C. canephora* e *C. eugenioides* serem espécies com alto grau de parentesco, é observada no híbrido uma baixa variação entre seus respectivos subgenomas [4]. Lashermes et al. relata a retenção e expressão de duas cópias de genes homeólogos em 96% dos genes [34], o que serve como uma possível explicação para o grande número de homólogos de diversos genes MIKC identificados, e resulta em uma redundância ainda maior para os integrantes da família MADS-box. Essas características são ilustradas na figura pela alta similaridade na distribuição de genes dentro de pares cromossômicos equivalentes, e também pelo número próximo de representantes da família MADS-box oriundos de cada subgenoma, onde, dos 87 genes posicionados, 43 e 44 são situados em cromossomos provenientes dos parentais *C. canephora* e *C. eugenioides*, respectivamente. Apesar disso, pequenas variações podem ser observadas, possivelmente como consequência de pequenos rearranjos genômicos envolvendo trocas homeólogas através de mecanismos de crossover ou conversão gênica [34].

De acordo com estudos similares em *Arabidopsis* [12] e em plantas de interesse econômico, como o arroz [15], a pereira [35] e o tomateiro [36], uma das características que distingue proteínas MADS-box tipo I e tipo II é o número de éxons presentes nos genes que as codificam. De maneira geral, genes tipo I são conhecidos por possuir apenas um éxon e um ou nenhum íntron, enquanto genes tipo II são mais estruturalmente complexos e contêm número variado de éxons e íntrons. Sabendo disso, realizamos uma análise estrutural dos genes identificados com o intuito de também confirmar as classificações atribuídas e investigar a ocorrência de variações nessa relação tipo-estrutura (Figura 4).

A partir da análise foi verificado que o número de íntrons nos genes MADS-box varia de 0 a 11, e pudemos confirmar a relação tipo-estrutura na maioria dos genes identificados. Entretanto, houveram algumas exceções. Por exemplo, foi verificada a presença de 3 íntrons nos genes *CaMADS26* e *CaMADS94*, e de 2 íntrons em *CaMADS100*, todos tipo I, em que esperaríamos um ou nenhum íntron, enquanto o gene MIKC* *CaMADS85* do tipo II não possui nenhum íntron quando múltiplos eram esperados. Além disso, é possível também observar uma alta similaridade estrutural entre genes proximalmente relacionados, ocorrendo variação apenas no comprimento de íntrons. Como exemplo temos os genes *CaMADS10*, *CaMADS13*, *CaMADS42* e *CaMADS48*, todos pertencentes ao subgrupo AGL17/AGL21 e de estrutura altamente semelhante, porém enquanto *CaMADS10* e *CaMADS13* possuem comprimento de 6-7 Kb, *CaMADS42* e *CaMADS48* possuem 25 Kb e 22 Kb, respectivamente.

Figura 4. Análise estrutural dos genes MADS-box do café arábico. As porções amarelas de cada gene representam éxons e as linhas contínuas, os íntrons. Na parte inferior da figura está situada a barra de escala, em Kilobases (Kb).



4. Considerações Finais

O presente estudo busca fornecer uma referência útil sobre os genes MADS-box e auxiliar na compreensão dos mecanismos moleculares responsáveis pelos processos dos desenvolvimentos vegetativo e reprodutivo do cafeeiro, contribuindo também na elucidação da história evolutiva desses importantes fatores de transcrição na família Rubiaceae. Nas próximas etapas determinaremos a expressão dos genes identificados em diferentes tecidos de *C. arabica* por meio de análises in silico de RNA-Seq, o que possibilitará uma análise mais aprofundada do grau de conservação desses genes e dos papéis exercidos por eles.

Apêndice A

Tabela A1. Relação dos 102 genes MADS-box identificados em *Coffea arabica*. Também informados na tabela estão seus respectivos identificadores no banco de dados GenBank, suas posições no genoma da planta, o grupo a que pertencem e o comprimento da proteína que codificam. * Proteínas descritas previamente por Oliveira et al. [8].

Gene name	Synonym	Genbank ID	Chr.	Localization	Group	Subgroup	Length (aa)
CaMADS1		XP_027062828.1	1E	31887001 - 31887848	Ma	NA	252
CaMADS2		XP_027063685.1	5C	40137420 - 40146426	MIKC	SHP	245
CaMADS3		XP_027064476.1	5C	1696972 - 1697820	Ma	NA	282
CaMADS4		XP_027068482.1	1E	32271015 - 32271721	Ma	NA	212
CaMADS5		XP_027069013.1	6C	15448302 - 15453697	MIKC	AGL15/ AGL18	240
CaMADS6		XP_027073038.1	6E	14187665 - 14192800	MIKC	AGL15/ AGL18	242
CaMADS7	CaTM6 *	XP_027073264.1	7C	7075691 - 7078254	MIKC	AP3	219
CaMADS8	CaSEP-1 *	XP_027074278.1	7C	2590468 - 2595678	MIKC	SEP	245
CaMADS9		XP_027075278.1	7C	10403072 - 10406815	MIKC	SVP	232
CaMADS10		XP_027076499.1	7C	5433796 - 5441026	MIKC	AGL17/ AGL21	235
CaMADS11		XP_027077910.1	7E	17073970 - 17086729	MIKC	AP1/FUL	242
CaMADS12		XP_027078034.1	7E	415170 - 420468	MIKC	SEP	245
CaMADS13	CaAGL17 *	XP_027078102.1	7E	3166861 - 3174028	MIKC	AGL17/ AGL21	235
CaMADS14	CaSVP-1 *	XP_027078661.1	7E	7604438 - 7608213	MIKC	SVP	232
CaMADS15		XP_027078843.1	7E	426275 - 434382	MIKC	AP1/FUL	241
CaMADS16	CaTM3-1 *	XP_027080181.1	8E	42234652 - 42244780	MIKC	AGL14/ AGL19	202
CaMADS17		XP_027082688.1	8E	44129029 - 44129646	Ma	NA	205
CaMADS18		XP_027082904.1	8C	36706443 - 36716138	MIKC	AGL14/ AGL19	202
CaMADS19		XP_027083082.1	8C	36538879 - 36548065	MIKC	AGL14/ AGL19	202
CaMADS20		XP_027083939.1	8C	38284598 - 38285213	Ma	NA	193
CaMADS21		XP_027083940.1	8C	38451695 - 38452312	Ma	NA	205
CaMADS22		XP_027085702.1	9C	7276705 - 7287809	MIKC	AGL14/ AGL19	213
CaMADS23		XP_027085703.1	9C	7235786 - 7252001	MIKC	AGL14/ AGL19	205
CaMADS24		XP_027086111.1	9C	7217661 - 7223289	MIKC	AGL14/ AGL19	212
CaMADS25		XP_027086639.1	9C	30079886 - 30080524	Ma	NA	212
CaMADS26		XP_027086640.1	9C	30169388 - 30173038	Ma	NA	321
CaMADS27		XP_027087702.1	9C	31053659 - 31055210	Ma	NA	374
CaMADS28		XP_027087965.1	9E	4010616 - 4011356	Ma	NA	181
CaMADS29	CaTM3-2 *	XP_027088367.1	9E	6409310 - 6414977	MIKC	AGL14/ AGL19	212
CaMADS30		XP_027088711.1	9E	29732528 - 29733652	Ma	NA	374

CaMADS31		XP_027088718.1	9E	29973413 - 29977164	Ma	NA	308
CaMADS32		XP_027088719.1	9E	30021212 - 30022336	Ma	NA	374
CaMADS33		XP_027089366.1	9E	6427515 - 6438609	MIKC	AGL14/ AGL19	204
CaMADS34		XP_027089367.1	9E	6472824 - 6489839	MIKC	AGL14/ AGL19	213
CaMADS35	CaPI *	XP_027089801.1	10E	36148862 - 36151534	MIKC	PI	210
CaMADS36		XP_027089990.1	10E	35757233 - 35761031	MIKC	TT16	254
CaMADS37		XP_027090010.1	10E	40052627 - 40061756	MIKC	SVP	227
CaMADS38		XP_027090646.1	10E	39980434 - 39989400	MIKC	SVP	227
CaMADS39	CaAP3 *	XP_027090983.1	10E	1625706 - 1628031	MIKC	AP3	224
CaMADS40		XP_027091062.1	10E	9111955 - 9117988	MIKC	AGL12	205
CaMADS41		XP_027091162.1	10E	9370046 - 9380330	MIKC	STK	224
CaMADS42	CaAGL16 *	XP_027092062.1	10E	7395733 - 7421886	MIKC	AGL17/ AGL21	239
CaMADS43		XP_027092769.1	10C	41014563 - 41017238	MIKC	PI	210
CaMADS44	CaSVP-2 *	XP_027092990.1	10C	44946055 - 44955263	MIKC	SVP	227
CaMADS45		XP_027093277.1	10C	40624080 - 40627376	MIKC	TT16	254
CaMADS46		XP_027093414.1	10C	7427346 - 7432528	MIKC	AGL12	205
CaMADS47		XP_027093918.1	10C	7720126 - 7730404	MIKC	STK	228
CaMADS48		XP_027094493.1	10C	6454180 - 6476418	MIKC	AGL17/ AGL21	239
CaMADS49	CaSEP-4 *	XP_027096631.1	11C	31815729 - 31822474	MIKC	SEP	239
CaMADS50	CaAP1 *	XP_027096636.1	11C	31806985 - 31812529	MIKC	AP1/FUL	242
CaMADS51	CaSEP-3 *	XP_027096722.1	11C	35639430 - 35644817	MIKC	SEP	246
CaMADS52		XP_027097746.1	11C	35614153 - 35629250	MIKC	FLC	206
CaMADS53		XP_027099549.1	11E	41742814 - 41747715	MIKC	SEP	241
CaMADS54	CaFLC *	XP_027100320.1	11E	41706090 - 41721807	MIKC	FLC	206
CaMADS55		XP_027101418.1	NA	5538 - 6263	Ma	NA	241
CaMADS56		XP_027101420.1	NA	21691 - 22425	Ma	NA	244
CaMADS57		XP_027101421.1	NA	27797 - 31962	Ma	NA	321
CaMADS58		XP_027101422.1	NA	44052 - 44723	Ma	NA	223
CaMADS59		XP_027101424.1	NA	50275 - 51797	Ma	NA	253
CaMADS60		XP_027101426.1	NA	65586 - 66257	Ma	NA	223
CaMADS61		XP_027101427.1	NA	69174 - 69839	Ma	NA	221
CaMADS62		XP_027101428.1	NA	81926 - 82597	Ma	NA	223
CaMADS63		XP_027101430.1	NA	477 - 1196	Ma	NA	240
CaMADS64		XP_027101434.1	NA	88133 - 89903	Ma	NA	240
CaMADS65		XP_027101456.1	NA	3463 - 4056	Ma	NA	197
CaMADS66		XP_027101457.1	NA	28132 - 28797	Ma	NA	221
CaMADS67	CaAG *	XP_027101517.1	2C	14611048 - 14618205	MIKC	AG	242
CaMADS68		XP_027101670.1	2C	17292698 - 17305368	MIKC	ANR1	240
CaMADS69		XP_027101746.1	NA	3523635 - 3524483	Ma	NA	282
CaMADS70		XP_027102172.1	NA	1773126 - 1773671	Ma	NA	181
CaMADS71		XP_027103881.1	2C	2838762 - 2844949	MIKC	SOC1	230
CaMADS72	CaSVP-3 *	XP_027106196.1	2C	17505089 - 17513253	MIKC	SVP	261
CaMADS73		XP_027108199.1	2C	63808808 - 63819347	MIKC	SOC1	216
CaMADS74		XP_027108862.1	2E	16923191 - 16923862	Ma	NA	223
CaMADS75	CaAGL6 *	XP_027109717.1	2E	68578009 - 68584569	MIKC	AGL6	253
CaMADS76		XP_027115031.1	2E	68835465 - 68842043	MIKC	AGL6	253

CaMADS77	CaAGL42 *	XP_027110410.1	2E	2974383 - 2981128	MIKC	SOC1	230
CaMADS78		XP_027112013.1	2E	13488581 - 13499653	MIKC	AG	242
CaMADS79		XP_027112766.1	2E	17272420 - 17284677	MIKC	ANR1	240
CaMADS80		XP_027113896.1	2E	29510281 - 29517123	MIKC	AGL15/ AGL18	311
CaMADS81	CaSOC1 *	XP_027115032.1	2E	68849447 - 68859978	MIKC	SOC1	216
CaMADS82		XP_027123772.1	1C	35773610 - 35774368	Ma	NA	252
CaMADS83	CaFUL *	XP_027125353.1	4E	23235158 - 23246543	MIKC	AP1/FUL	242
CaMADS84		XP_027123666.1	4C	38954926 - 38962701	MIKC	AP1/FUL	423
CaMADS85		XP_027101419.1	NA	10599 - 11324	MIKC*	NA	241
CaMADS86	CaMD *	XP_027125662.1	4E	39220405 - 39227948	MIKC*	NA	409
CaMADS87		XP_027062683.1	5C	2878627 - 2879736	Mg	NA	369
CaMADS88		XP_027093617.1	10C	19156892 - 19157998	Mg	NA	368
CaMADS89		XP_027078767.1	7E	5440936 - 5441622	Mg	NA	228
CaMADS90		XP_027109449.1	2E	62460168 - 62461394	Mg	NA	408
CaMADS91		XP_027062843.1	1E	32180000 - 32180799	Ma	NA	236
CaMADS92		XP_027075722.1	7C	11966760 - 11969493	MIKC*	NA	232
CaMADS93		XP_027076830.1	7E	9174844 - 9177730	MIKC*	NA	232
CaMADS94		XP_027086465.1	9C	7596121 - 7597179	Ma	NA	246
CaMADS95		XP_027087159.1	9C	3390213 - 3392532	Mg	NA	320
CaMADS96		XP_027087795.1	9E	2898683 - 2900906	Mg	NA	321
CaMADS97		XP_027095848.1	11C	24509371 - 24510275	Mg	NA	224
CaMADS98		XP_027098569.1	11E	31065295 - 31066206	Mg	NA	224
CaMADS99		XP_027102546.1	2C	62725975 - 62726754	Mg	NA	259
CaMADS100		XP_027066504.1	6C	3395314 - 3396479	Mg	NA	189
CaMADS101		XP_027091235.1	10E	4307971 - 4311424	MIKC*	NA	329
CaMADS102		XP_027075649.1	7C	8298062 - 8298748	Mg	NA	228

Referências

- Davis, A.P.; Govaerts, R.; Bridson, D.M.; Stoffelen, P. An annotated taxonomic conspectus of the genus *Coffea* (Rubiaceae). *Bot. J. Linn. Soc.* **2006**, *152*, 465–512, doi:10.1111/j.1095-8339.2006.00584.x.
- Hein, L.; Gatzweiler, F. The economic value of coffee (*Coffea arabica*) genetic resources. *Ecol. Econ.* **2006**, *60*, 176–185, doi:10.1016/j.ecolecon.2005.11.022.
- Lashermes, P.; Combes, M.C.; Robert, J.; Trouslot, P.; D'Hont, A.; Anthony, F.; Charrier, A. Molecular characterisation and origin of the *Coffea arabica* L. Genome. *Mol. Gen. Genet.* **1999**, *261*, 259–266, doi:10.1007/s004380050965.
- Scalabrin, S.; Toniutti, L.; Di Gaspero, G.; Scaglione, D.; Magris, G.; Vidotto, M.; Pinosio, S.; Cattonaro, F.; Magni, F.; Jurman, I.; et al. A single polyploidization event at the origin of the tetraploid genome of *Coffea arabica* is responsible for the extremely low genetic variation in wild and cultivated germplasm. *Sci. Rep.* **2020**, *10*, 1–13, doi:10.1038/s41598-020-61216-7.
- Paes, Â.; Camargo, D.E.; Bento, M.; De Camargo, P. AGROMETEOROLOGIA DEFINIÇÃO E ESQUEMATIZAÇÃO DAS FASES FENOLÓGICAS DO CAFEIEIRO ARÁBICA NAS CONDIÇÕES TROPICAIS DO BRASIL (1).

6. Coen, E.S.; Meyerowitz, E.M. The war of the whorls: Genetic interactions controlling flower development. *Nature* 1991, 353, 31–37.
7. Gramzow, L.; Theissen, G. A hitchhiker's guide to the MADS world of plants. *Genome Biol.* **2010**, *11*, 1–11, doi:10.1186/gb-2010-11-6-214.
8. Schwarz-Sommer, Z.; Huijser, P.; Nacken, W.; Saedler, H.; Sommer, H. Genetic control of flower development by homeotic genes in *Antirrhinum majus*. *Science (80-.)*. **1990**, *250*, 931–936, doi:10.1126/science.250.4983.931.
9. de Oliveira, R.R.; Chalfun-Junior, A.; Paiva, L.V.; Andrade, A.C. In silico and quantitative analyses of MADS-box genes in *Coffea arabica*. *Plant Mol. Biol. Report.* **2010**, *28*, 460–472, doi:10.1007/s11105-009-0173-5.
10. de Oliveira, R.R.; Cesarino, I.; Mazzafera, P.; Dornelas, M.C. Flower development in *Coffea arabica* L.: New insights into MADS-box genes. *Plant Reprod.* **2014**, *27*, 79–94, doi:10.1007/s00497-014-0242-2.
11. Barreto, H.G.; Lazzari, F.; Ságio, S.A.; Chalfun-Junior, A.; Paiva, L.V.; Benedito, V.A. In Silico and Quantitative Analyses of the Putative FLC-like Homologue in Coffee (*Coffea arabica* L.). *Plant Mol. Biol. Report.* **2012**, *30*, 29–35, doi:10.1007/s11105-011-0310-9.
12. Pařenicová, L.; De Folter, S.; Kieffer, M.; Horner, D.S.; Favalli, C.; Busscher, J.; Cook, H.E.; Ingram, R.M.; Kater, M.M.; Davies, B.; et al. Molecular and phylogenetic analyses of the complete MADS-Box transcription factor family in *Arabidopsis*: New openings to the MADS world. *Plant Cell* **2003**, *15*, 1538–1551, doi:10.1105/tpc.011544.
13. Ng, M.; Yanofsky, M.F. Function and evolution of the plant MADS-box gene family. *Nat. Rev. Genet.* **2001**, *2*, 186–195, doi:10.1038/35056041.
14. Kofuji, R.; Sumikawa, N.; Yamasaki, M.; Kondo, K.; Ueda, K.; Ito, M.; Hasebe, M. Evolution and Divergence of the MADS-Box Gene Family Based on Genome-Wide Expression Analyses. *Mol. Biol. Evol.* **2003**, *20*, 1963–1977, doi:10.1093/molbev/msg216.
15. Arora, R.; Agarwal, P.; Ray, S.; Singh, A.K.; Singh, V.P.; Tyagi, A.K.; Kapoor, S. MADS-box gene family in rice: Genome-wide identification, organization and expression profiling during reproductive development and stress. *BMC Genomics* **2007**, *8*, 1–21, doi:10.1186/1471-2164-8-242.
16. Masiero, S.; Colombo, L.; Grini, P.E.; Schnittger, A.; Kater, M.M. The emerging importance of type I MADS box transcription factors for plant reproduction. *Plant Cell* **2011**, *23*, 865–872, doi:10.1105/tpc.110.081737.
17. Leseberg, C.H.; Li, A.; Kang, H.; Duvall, M.; Mao, L. Genome-wide analysis of the MADS-box gene family in *Populus trichocarpa*. *Gene* **2006**, *378*, 84–94, doi:10.1016/j.gene.2006.05.022.
18. Theissen, G. Development of floral organ identity: Stories from the MADS house. *Curr. Opin. Plant Biol.* **2001**, *4*, 75–85, doi:10.1016/S1369-5266(00)00139-4.

-
19. Krizek, B.A.; Fletcher, J.C. Molecular mechanisms of flower development: An armchair guide. *Nat. Rev. Genet.* **2005**, *6*, 688–698.
 20. Favaro, R.; Pinyopich, A.; Battaglia, R.; Kooiker, M.; Borghi, L.; Ditta, G.; Yanofsky, M.F.; Kater, M.M.; Colombo, L. MADS-Box Protein Complexes Control Carpel and Ovule Development in Arabidopsis. *Plant Cell* **2003**, *15*, 2603–2611, doi:10.1105/tpc.015123.
 21. Brambilla, V.; Battaglia, R.; Colombo, M.; Masiero, S.; Bencivenga, S.; Kater, M.M.; Colombo, L. Genetic and molecular interactions between BELL1 and MADS box factors support ovule development in arabidopsis. *Plant Cell* **2007**, *19*, 2544–2556, doi:10.1105/tpc.107.051797.
 22. Immink, R.G.H.; Tonaco, I.A.N.; de Folter, S.; Shchennikova, A.; van Dijk, A.D.J.; Busscher-Lange, J.; Borst, J.W.; Angenent, G.C. SEPALLATA3: The “glue” for MADS box transcription factor complex formation. *Genome Biol.* **2009**, *10*, 24, doi:10.1186/gb-2009-10-2-r24.
 23. Berardini, T.Z.; Reiser, L.; Li, D.; Mezheritsky, Y.; Muller, R.; Strait, E.; Huala, E. The arabidopsis information resource: Making and mining the “gold standard” annotated reference plant genome. *Genesis* **2015**, *53*, 474–485, doi:10.1002/dvg.22877.
 24. Altschul, S.F.; Gish, W.; Miller, W.; Myers, E.W.; Lipman, D.J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–410, doi:10.1016/S0022-2836(05)80360-2.
 25. Mistry, J.; Chuguransky, S.; Williams, L.; Qureshi, M.; Salazar, G.A.; Sonnhammer, E.L.L.; Tosatto, S.C.E.; Paladin, L.; Raj, S.; Richardson, L.J.; et al. Pfam: The protein families database in 2021. *Nucleic Acids Res.* **2021**, *49*, D412–D419, doi:10.1093/nar/gkaa913.
 26. Letunic, I.; Khedkar, S.; Bork, P. SMART: Recent updates, new developments and status in 2020. *Nucleic Acids Res.* **2021**, *49*, D458–D460, doi:10.1093/nar/gkaa937.
 27. Rozewicki, J.; Li, S.; Amada, K.M.; Standley, D.M.; Katoh, K. MAFFT-DASH: Integrated protein sequence and structural alignment. *Nucleic Acids Res.* **2019**, *47*, W5–W10, doi:10.1093/nar/gkz342.
 28. Sela, I.; Ashkenazy, H.; Katoh, K.; Pupko, T. GUIDANCE2: Accurate detection of unreliable alignment regions accounting for the uncertainty of multiple parameters. *Nucleic Acids Res.* **2015**, *43*, W7–W14, doi:10.1093/nar/gkv318.
 29. Felsenstein, J. PHYLIP - Phylogeny Inference Package (Ver. 3.2). *Cladistics* **1989**, *5*, 164–166.
 30. Jiangtao, C.; Yingzhen, K.; Qian, W.; Yuhe, S.; Daping, G.; Jing, L.; Guanshan, L. MapGene2Chrom, a tool to draw gene physical map based on Perl and SVG languages. *Yi Chuan* **2015**, *37*, 91–97, doi:10.16288/j.ycz.2015.01.013.
 31. Hu, B.; Jin, J.; Guo, A.Y.; Zhang, H.; Luo, J.; Gao, G. GSDS 2.0: An upgraded gene feature visualization server. *Bioinformatics* **2015**, *31*, 1296–1297, doi:10.1093/bioinformatics/btu817.

-
32. Grimplet, J.; Martínez-Zapater, J.M.; Carmona, M.J. Structural and functional annotation of the MADS-box transcription factor family in grapevine. *BMC Genomics* **2016**, *17*, 1–23, doi:10.1186/s12864-016-2398-7.
 33. Zhang, H.; Forde, B.G. An Arabidopsis MADS box gene that controls nutrient-induced changes in root architecture. *Science* (80-.). **1998**, *279*, 407–409, doi:10.1126/science.279.5349.407.
 34. Lashermes, P.; Hueber, Y.; Combes, M.C.; Severac, D.; Dereeper, A. Inter-genomic DNA exchanges and homeologous gene silencing shaped the nascent allopolyploid coffee genome (*Coffea arabica* L.). *G3 Genes, Genomes, Genet.* **2016**, *6*, 2937–2948, doi:10.1534/g3.116.030858.
 35. Meng, D.; Cao, Y.; Chen, T.; Abdullah, M.; Jin, Q.; Fan, H.; Lin, Y.; Cai, Y. Evolution and functional divergence of MADS-box genes in *Pyrus*. *Sci. Rep.* **2019**, *9*, 1–13, doi:10.1038/s41598-018-37897-6.
 36. Wang, Y.; Zhang, J.; Hu, Z.; Guo, X.; Tian, S.; Chen, G. Genome-wide analysis of the MADS-box transcription factor family in *Solanum lycopersicum*. *Int. J. Mol. Sci.* **2019**, *20*, 2961, doi:10.3390/ijms20122961.