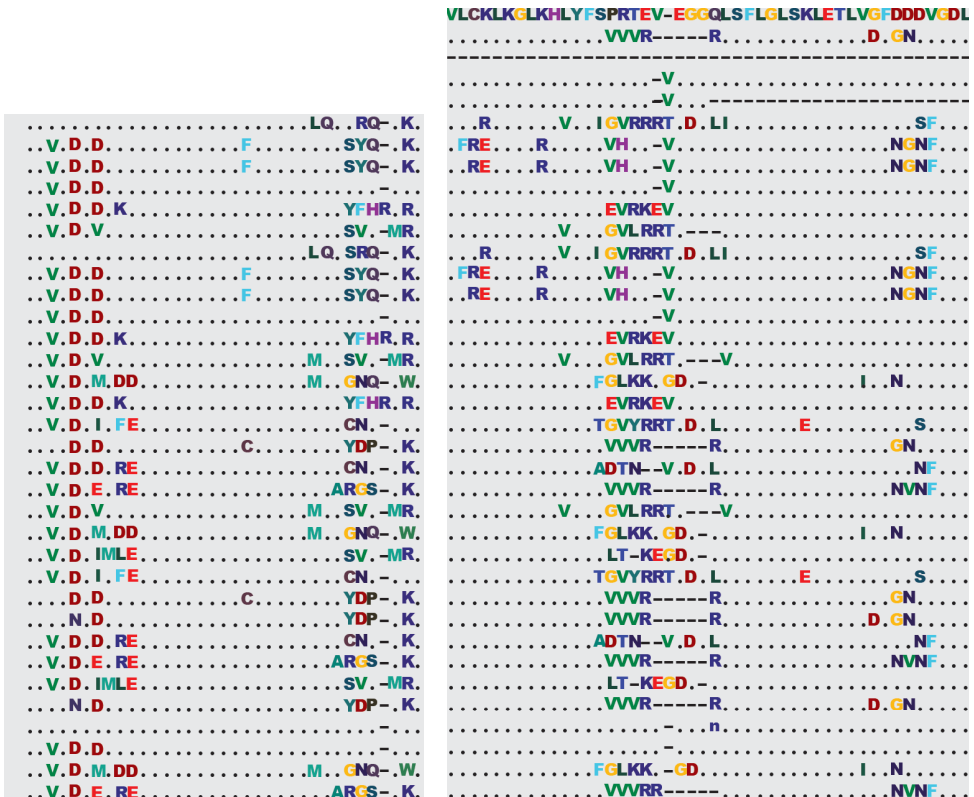


Os Loci S_H3 Envolvidos na Resistência à Ferrugem são Complexos, Multialélicos e Divergentes em Genomas de *Coffea*



**Empresa Brasileira de Pesquisa Agropecuária
Embrapa Café
Ministério da Agricultura, Pecuária e Abastecimento**

**BOLETIM DE PESQUISA
E DESENVOLVIMENTO
3**

Os Loci S_H3 Envolvidos na Resistência à
Ferrugem são Complexos, Multialélicos e
Divergentes em Genomas de *Coffea*

*Paula Cristina da Silva Angelo
Eveline Teixeira Caixeta
Luiz Filipe Protásio Pereira
Alessandra F. Ribas
Gustavo H. Sera*

Embrapa Café
Brasília, DF
2022

Exemplares desta publicação podem ser adquiridos na:

Embrapa Café
Parque Estação Biológica (PqEB)
Av. W3 Norte (final), Ed. Sede
CEP: 70770-901, Brasília, DF
Fone: +55 (61) 3448-4378 / 4010
Fax: +55 (61) 3448-1797
www.embrapa.br
www.embrapa.br/fale-conosco/sac

Comitê Local de Publicações da Embrapa Café

Presidente
Lucas Tadeu Ferreira

Vice-Presidente
Jamilsen de Freitas Santos

Secretária-Executiva
Adriana Maria Silva Macedo

Membros
Anísio José Diniz, Carlos Henrique Siqueira de Carvalho, Helena Maria Ramos Alves, Lucilene Maria de Andrade, Mauricio Sergio Zacarias, Milene Alves de Figueiredo Carvalho, Omar Cruz Rocha, Rogério Novais Teixeira, Roseane Pereira Villela.

Revisão de texto
Ana Maranhao Nogueira

Normalização bibliográfica
Maria de Fátima da Cunha

Tratamento das ilustrações
Thiago Farah Cavaton

Projeto gráfico da coleção
Carlos Eduardo Felice Barbeiro

Editoração eletrônica
Thiago Farah Cavaton

Foto da capa
Paula Cristina da Silva Angelo

1ª edição
Publicação digital (2022)

Todos os direitos reservados.

A reprodução não autorizada desta publicação, no todo ou em parte, constitui violação dos direitos autorais (Lei nº 9.610).

Dados Internacionais de Catalogação na Publicação (CIP)

Embrapa, Secretaria-Geral

Os LOCI S_H3 envolvidos na resistência à ferrugem são complexos multialélicos e divergentes em genomas de *Coffea* / Paula Cristina da Silva Angelo ... [et al.]. – Brasília, DF : Embrapa Café, 2022.
PDF (19 p.). – (Boletim de Pesquisa e Desenvolvimento / Embrapa Café, ISSN 2237-9738 ; 3).

1. Café. 2. *Coffea arabica*. 3. Ferrugem alaranjada. 3. *Hemileia vastatrix*. 4. Polimorfismo. 5. Genoma. I. Angelo, Paula Cristina da Silva. II. Caixeta, Eveline Teixeira. III. Pereira, Luiz Filipe Protásio. IV. Ribas, Alessandra Ferreira. V. Sera, Gustavo Hiroshi. VI. Série.

CDD 633.73

Sumário

Resumo	5
Abstract	6
Introdução.....	7
Material e Métodos	10
Resultados e Discussão	11
Conclusões.....	18
Referências	18

Os Loci S_H3 Envolvidos na Resistência à Ferrugem são Complexos, Multialélicos e Divergentes em Genomas de *Coffea*¹

Paula Cristina da Silva Angelo²

Eveline Teixeira Caixeta³

Luiz Filipe Protásio Pereira⁴

Alessandra F. Ribas⁵

Gustavo H. Sera⁶

Resumo – Os genes no locus S_H3 são importantes para resistência duradoura dos cafeeiros à ferrugem, causada pelo fungo *Hemileia vastatrix*. Há um agrupamento de oito genes que codificam proteínas R da família CC-NBS-LRR no locus S_H3 da variedade IAPAR59 de *Coffea arabica*. A sequência dessas proteínas R é variável, especialmente na região carboxiterminal, onde estão localizados motivos ricos em leucina (LRR) que participam do reconhecimento das diferentes raças do fungo. Os objetivos do trabalho foram investigar: 1) a estrutura do locus S_H3 em *C. arabica* cv. Caturra e silvestre de origem etíope, em *C. canephora* DH200-94 e em *C. eugenoides* que tiveram os genomas sequenciados; e 2) a variabilidade da região carboxiterminal dos peptídeos deduzidos para todas as cópias dos genes R em loci S_H3 desses quatro genótipos e de IAPAR59 e *C. canephora* IF200. Os peptídeos foram identificados, alinhados e agrupados pelo método de Fitch-Margoliash. Concluiu-se que o locus S_H3 é complexo e multialélico em todos os genomas investigados. Há diferença no número de cópias e na organização e na sequência das cópias dos genes R nos loci S_H3 dos diferentes genomas, características que podem estar associadas ao mecanismo de distinção e reconhecimento das raças fisiológicas de *H. vastatrix* pelas variedades de cafeeiros.

Termos para indexação: *Coffea* spp., Doença do cafeeiro, Polimorfismo estrutural, Isoformas protéicas.

¹ Trabalho financiado pelo Consórcio Brasileiro de Pesquisa e Desenvolvimento do Café (Consórcio Pesquisa Café).

² Bióloga, D.Sc. em Ciências Biológicas, ênfase em Genética Vegetal, pesquisadora da Embrapa Café/Instituto de Desenvolvimento Rural do Paraná, Londrina, PR

³ Engenheira-agrônoma, Ph.D. em Genética e Melhoramento, pesquisadora da Embrapa Café/ Universidade Federal de Viçosa, Viçosa, MG

⁴ Biólogo, Ph.D. em Genética Vegetal, pesquisador da Embrapa Café/ Instituto de Desenvolvimento Rural do Paraná, Londrina, PR

⁵ Agrônoma, Ph.D. em Biologia Integrada de Plantas, professora da Universidade do Oeste Paulista, Presidente Prudente, SP

⁶ Agrônomo, Ph.D. em Melhoramento e Genética de Plantas, pesquisador do Instituto de Desenvolvimento Rural do Paraná, Londrina, PR

S_H3 Loci Involved in the Resistance to Leaf Rust are Complex, Multi Allelic and Divergent in *Coffea*

Abstract – The genes in the S_H3 locus are important for the durable resistance to the coffee leaf rust disease caused by the fungi *Hemileia vastatrix*. There is a cluster of eight genes coding for CC-NBS-LRR R proteins in the S_H3 locus of *Coffea arabica* cv. IAPAR59. The R protein sequences, specially the carboxy-terminal region where leucine-rich-repeats (LRR) are placed to take their roles in recognizing the different fungi races, are variable. The objectives of this work were to investigate: 1) the S_H3 structure in *C. arabica* cv. Caturra and native of Ethiopia, in *C. canephora* DH200-94 and *C. eugenioides*, which had their genomes recently sequenced; and 2) the variability in the carboxy-terminal region of peptides deduced from R gene copies from those four genotypes and *C. arabica* cv. IAPAR59 and *C. canephora* IF200. The peptides were identified, aligned and clustered by the Fitch-Margoliash method. It was concluded that the S_H3 loci are complex and multi-allelic in all the investigated genomes. There is a different number of R gene copies displaying different sequences, set in different positions and directions in S_H3 loci of the investigated genomes, and these characteristics can be related to the coffee mechanisms to distinguish and recognize the different *H. vastatrix* physiological races.

Index terms: *Coffea* spp., Coffee disease, Structural polymorphism, Protein isoforms.

Introdução

A ferrugem, causada pelo fungo biotrófico *Hemileia vastatrix* Berkeley & Broome, é uma das principais doenças do cafeeiro arábica, pois reduz a produtividade devido à desfolha das plantas e à morte de ramos (Matiello et al., 2016; Zambolim, 2016). A interação entre o fungo e o cafeeiro pode ser classificada como interação do tipo “gene-a-gene” (Talhinhas et al., 2017), que ocorre quando uma proteína R do cafeeiro, codificada em um gene de resistência (gene R), reconhece especificamente um efector do fungo (Barka et al., 2017). Quando é reconhecido pelo cafeeiro, o efector é classificado como fator de avirulência ou *avr*, as reações de defesa do cafeeiro são então disparadas e os sintomas graves da ferrugem são bloqueados. Quando nenhuma proteína R reconhece os efetores de uma determinada raça de *H. vastatrix*, reações de defesa do tipo gene-a-gene deixam de ser ativadas e/ou são ativadas tardiamente, o que permite que o fungo se instale e que os sintomas da ferrugem se desenvolvam no cafeeiro. Alterações na sequência dos genes de resistência das variedades de cafeeiros estão relacionadas com alterações na sequência dos efetores presentes nas diferentes raças de *H. vastatrix*. Por seleção natural, essa relação vai resultar ora no aprimoramento da resistência de uma variedade de cafeeiro, ora no aprimoramento da virulência e capacidade de causar doença de uma raça de fungo (Flor, 1971).

A ocorrência de alterações simultâneas em muitos genes relacionados com qualquer característica fenotípica é menos frequente do que a alteração em apenas um gene por geração, assim, cafeeiros que expressam vários genes de resistência, quando são expostos aos fungos, têm mais chance de reconhecê-los, ativar as defesas e continuarem produtivos por longos períodos. Esse fato é especialmente importante, pois o número de propágulos produzidos pelos fungos é muito maior do que o número de plântulas com genótipos novos produzidos por cultivares de cafeeiros considerando-se um mesmo intervalo de tempo. Os fungos, portanto, estão muito mais propensos a sofrer e propagar alterações genéticas do que os cafeeiros ao longo do tempo. Além disso, *H. vastatrix* conta com um mecanismo descrito recentemente de sexualidade críptica, que aumenta a probabilidade do aparecimento de novos efetores por recombinação entre genes de efetores pré-existentes (Carvalho et al., 2011). Por fim, os fungos acessam rapidamente diferentes localidades, por exemplo, quando os esporos, que são produzidos aos milhões nos cam-

pos, são carregados pelo vento, em roupas ou em veículos (Schieber, 1972) e, assim, podem infectar cafeeiros suscetíveis. Quando o ambiente e o hospedeiro lhes são favoráveis, os fungos colonizam uma área tão extensa dos tecidos foliares dos cafeeiros, que provocam a queda de folhas de grande número de ramos inteiros (Schieber, 1972; Matiello et al., 2016).

Os produtos dos genes R dos loci S_H1 a S_H9 dos cafeeiros são proteínas de resistência à ferrugem. Dessa forma, o desenvolvimento de cultivares com vários genes S_H que codifiquem diferentes proteínas de resistência é de extrema importância para obter uma resistência durável (Sera et al., 2010). Ademais, foram registradas no Brasil, pelo menos, 15 raças de *H. vastatrix* (Capucho et al., 2012; Zambolim, 2016), entre elas a raça XXXVII, que apresenta genes de virulência *v2*, *5*, *6*, *7* e *9*, divulgada em artigo científico em 2009, a qual, à época, poderia ter infectado qualquer planta cultivada comercialmente no País (Cabral et al., 2009).

A introgressão de genes S_H3 de *Coffea liberica* em *C. arabica* é, então, buscada há várias décadas para conferir resistência a um maior número de raças de *H. vastatrix* (Carvalho et al., 1977; Sera et al., 2007; Mahé et al., 2008; Alkimim et al., 2017). Essa resistência seria duradoura, podendo chegar a mais de 50 anos, segundo observações dos Pesquisadores do Instituto Agrônomo (IAC) (Conceição et al., 2005). A Biologia Molecular, por sua vez, pode contribuir com a análise da sequência e da estrutura dos genes de resistência; com o desenvolvimento de marcadores moleculares que assistirão o trabalho de seleção precoce de plantas com a configuração genética ideal ou próxima do ideal (entre as milhares que são produzidas ao longo dos programas de melhoramento); e, potencialmente, com o desenvolvimento de plantas com genes de resistência introduzidos por transgenia. Pesquisas em que as ferramentas da Biologia Molecular são utilizadas no melhoramento são oportunidades para a geração e disponibilização para a sociedade de cultivares com muito valor agregado em tempo relativamente curto, o que é uma vantagem no controle de doenças que se propagam com relativa facilidade.

O locus S_H3 de *C. arabica* (espécie oriunda de hibridização espontânea e interespecífica entre *C. eugenoides* x *C. canephora* ocorrida há milhões de anos) variedade cultivada IAPAR59 foi clonado e sequenciado por Ribas et al. (2011). É uma região com cerca de 250.000 pares de bases, onde existe um agrupamento de oito genes, estando cinco deles em cromossomas oriundos do ancestral *C. eugenoides* e três em cromossomas oriundos do ancestral *C. canephora*.

Cada um dos genes que compõem o agrupamento S_H3 codifica uma proteína da família CC-NBS-LRR. As CC-NBS-LRRs são assim denominadas porque apresentam um domínio amino-terminal com estrutura terciária em forma de *coiled-coil* (CC), um domínio central de ligação a nucleotídeos (*nucleotide binding sequence* – NBS) e um domínio carboxiterminal que contém número variável de motivos com frequência alta do aminoácido leucina, denominados *leucine rich repeats* (LRR). Esse último domínio é o que apresenta maior variabilidade na sequência de aminoácidos dessas proteínas de resistência e é também o domínio mais envolvido na interação e reconhecimento dos efetores dos patógenos (ou outras proteínas da célula que são expressas durante as reações de defesa). O histórico de duplicações, *crossing over* desigual, recombinação ectópica, conversão gênica e diversificação induzida por seleção que acompanha as proteínas de resistência CC-NBS-LRRs em diversas espécies vegetais além do cafeeiro, é atribuído à necessidade de interagir com patógenos que têm ciclo de vida reduzido e grande capacidade de produzir efetores alternativos, conforme já explicado. O agrupamento de várias CC-NBS-LRR em um mesmo locus também não é exclusividade dos cafeeiros (Nepal et al., 2017). A complexidade dessas interações de reconhecimento e defesa entre planta e patógeno é bastante reconhecida e estudada (Bonardi; Dangl, 2012).

No entanto, apesar de possuir oito genes que codificam proteínas CC-NBS-LRRs em seu locus S_H3 , pelo menos seis deles completamente funcionais (Ribas et al., 2011), a IAPAR59 é uma cultivar com fenótipo S_H3^- . Por outro lado, a introgressão de genes S_H3 provenientes de IAC 1110 (cafeeiro BA-10 introduzido da Índia) em cafeeiros Catuaí determinou, em discos de folhas das plantas selecionadas após os cruzamentos IAC 1110 x Catuaí, a ocorrência do fenótipo de resistência à inoculação de fungos das raças I, II e III, que têm os efetores $v2+v5$, $v5$ e $v1+v5$, respectivamente (Conceição et al., 2005). Essas informações permitem inferir a complexidade do funcionamento do S_H3 .

Para o presente trabalho, os agrupamentos de genes de resistência da família CC-NBS-LRR em loci homólogos ao S_H3 identificado por Ribas et al. (2011) no IAPAR59 foram identificados nos genomas sequenciados recentemente de *C. arabica*, *C. canephora* e *C. eugenioides*. Tendo identificado homólogos do S_H3 em quatro genomas, verificou-se que há diversidade quanto ao número e organização dos genes nesse locus e que a estrutura primária do domínio LRR das proteínas CC-NBS-LRR, que são produtos desses genes, diverge tanto dentro de cada genoma quanto entre diferentes genomas.

Material e Métodos

As sequências de nucleotídeos, incluindo os íntrons, que codificam os polipeptídeos publicados por Ribas et al. (2011), foram utilizadas para buscar os homólogos nos genomas de *C. arabica*, *C. canephora* e *C. eugenioides*. A partir das sequências de DNA genômico recuperadas (ANGELO, 2021), foram deduzidos os peptídeos codificados pela região carboxiterminal (596 resíduos de aminoácidos da região carboxiterminal, incluindo o domínio LRR) de cada cópia dos genes codificadores de CC-NBS-LRRs encontrada nos agrupamentos $S_{H,3}$ dos diferentes genomas. Essas regiões foram alinhadas e comparadas utilizando o algoritmo de Fitch-Margoliash (taxa de substituição livre ao longo de toda a sequência; version 3.5c, Phylogeny Inference Package – PHYLIP⁷) incluído no aplicativo BIOEDIT v.7.2.6 (Hall, 1999). Os bioprojetos consultados no National Center for Biotechnology Information (NCBI) até junho de 2019⁸ foram *C. arabica* cv. Caturra Vermelho Cenicafé (tetraplóide), PRJNA 497895 sequências referência e não referência; *C. eugenioides* (diplóide) acesso CCC68 do Cenicafé, Colômbia, PRJNA 508372 e PRJNA 497891 sequências referência e não referência; *C. canephora* diplóide DH200-94 IRD França, PRJEB 4211. As sequências desse mesmo genótipo de *C. canephora* depositadas no Coffee Hub⁹ também foram acessadas. Por fim, foram acessadas sequências do genoma de um indivíduo de *C. arabica* (tetraplóide), designado como *CaOE* ao longo do presente trabalho, parte de população silvestre natural da Etiópia, mantida ex situ no Instituto de Desenvolvimento Rural do Paraná (IDR-Paraná) (previamente denominado Iapar), em Londrina. O sequenciamento foi realizado pelo Consórcio Genoma Café Arábica (ACGC). As sequências de aminoácidos analisadas ao longo deste trabalho foram codificadas como: **ARA** para sequências de *C. arabica*, depositadas no **NCBI** pela Johns Hopkins University (**JH**), classificadas como *reference sequences* (**RF**) ou não; **ARANP** para *CaOE*, *C. arabica* silvestre de origem etíope; **CAN** para sequências de *C. canephora*, depositadas no **NCBI** ou no Coffee Hub; e **EU** para sequências de *C. eugenioides*, depositadas no **NCBI** pela Johns Hopkins University, classificadas como *reference sequences* ou não.

⁷ Disponível em: <http://evolution.gs.washington.edu/phylip.html>

⁸ Disponível em: www.ncbi.nlm.nih.gov

⁹ Disponível em: www.coffeehub.org

Resultados e Discussão

Em todos os genomas acessados, as regiões homólogas ao S_H3 de *C. arabica* cv. IAPAR59 sequenciado por Ribas et al. (2011) foram localizadas em cromossomas de número 3. Diferentes números de cópias dos genes codificando proteínas CC-NBS-LRR reunidas no S_H3 foram encontrados nos diferentes genomas.

Em *C. arabica* silvestre de origem etíope (*CaOE*) foi observado que o agrupamento apresenta organização semelhante à registrada para a cultivar IAPAR59 em Ribas et al. (2011) mas há dois genes a menos em *CaOE*. Portanto, o S_H3 em *CaOE* é formado por dois genes CC-NBS-LRR em cromossomas no 3 oriundos de *C. canephora* (A^C) e outros quatro em cromossomas no 3 oriundos de *C. eugenioides* (A^E) (Figura 1). A orientação dos genes e o número e tamanho dos espaçadores entre eles é de 27,3 kb em A^C e de 168,4, 14,8 e 26,9 kb em A^E , o que é muito semelhante ao que foi descrito para IAPAR59. A estrutura das diferentes cópias do gene inseridas no agrupamento também é bastante preservada, com o exon 1 variando de 1.040 a 1.045 pb, íntrons em torno de 250 pb e o exon 2 apresentando entre 1.773 e 1.795 pb. Uma das cópias do gene, *ARANP5* (Figura 1, seta 5) do complemento A^E em *CaOE*, tem, aparentemente, um códon de terminação precoce e pode ser um pseudogene, à semelhança do que foi observado por Ribas et al. (2011) para as cópias B2-Ea e A2-Ea do IAPAR59.

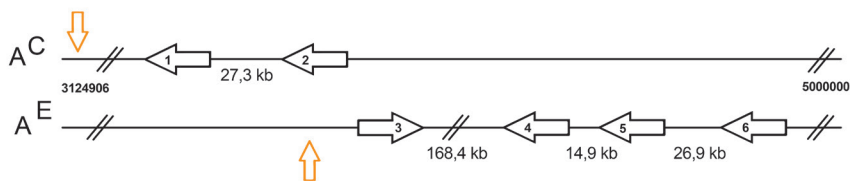


Figura 1. Organização do agrupamento de genes de resistência à ferrugem no locus S_H3 de *Coffea arabica* silvestre de origem etíope (*CaOE*). As setas pretas indicam a orientação e a localização de cada uma das seis cópias de regiões codificadoras de proteínas R da família CC-NBS-LRR. As setas vermelhas indicam a posição de duas cópias do gene que foram observadas na cultivar IAPAR59 (Ribas et al., 2011), mas que estão ausentes do genoma de *CaOE*. Os números nas extremidades da figura indicam pares de base. A^C = cromossoma 3 do complemento oriundo de *C. canephora*. A^E = cromossoma 3 do complemento oriundo de *C. eugenioides*. A representação não está em escala, portanto as posições são aproximadas. Os números de 1 a 6 dentro das setas são simples indicações da ordem em que os genes foram encontrados nos cromossomas, de acordo com a montagem disponível até junho de 2019.

No cromossoma 3 de *C. eugenoides* CCC68, cujo genoma está disponível para a consulta no site do NCBI, foram encontradas quatro cópias de genes *R* e espaçadores de 219, 15 e 16 kb, aproximadamente. Em *C. canephora* DH200-94, foram encontradas quatro cópias no cromossoma 3, espaçadas por 226, 16 e 11 kb. Em cada um desses dois últimos genomas, uma cópia isolada do gene foi localizada nos cromossomas 6, portanto fora dos agrupamentos S_{H3} onde as demais cópias se organizaram. Em *C. arabica* cv. Caturra, foram encontradas quatro cópias do gene, com espaçadores de 7,5, 182 e 19 kb entre cópias, todas no complemento genômico oriundo de *C. canephora*. Uma das quatro cópias (*ARAJH4*) parece ter um códon de terminação precoce e pode tratar-se de um pseudogene.

A diversidade da sequência das regiões carboxiterminais, incluindo os domínios LRR (596 aminoácidos em peptídeos deduzidos da sequência dos genes) de 44 peptídeos CC-NBS-LRRs foi avaliada. Essas regiões (ou domínios) foram escolhidas para a avaliação por sua importância, já mencionada, para a especificidade da interação dos cafeeiros com as diferentes raças de *H. vastatrix* (Ribas et al., 2011; Nepal et al., 2017). Na avaliação, foram incluídas todas as cópias de peptídeos reunidas nos loci S_{H3} dos cromossomas 3, assim como aquelas nos cromossomas 6, de todos os genótipos citados das espécies *C. arabica*, *C. canephora* e *C. eugenoides* disponibilizadas para consulta on-line, ou de *CaOE*. Também foram incluídos na avaliação de diversidade os peptídeos deduzidos das oito cópias de genes *R* identificadas no locus S_{H3} do genoma de IAPAR59 (contando com as cópias A2-E^a e B2-E^a, que apresentam códons de terminação precoces) e das quatro cópias de genes *R* identificadas no genoma de *C. canephora* IF200 por Ribas et al. (2011). As sequências designadas como de referência (curadas pelo NCBI) e as sequências não referência (também depositadas no NCBI) dos genomas de *C. arabica* var. Caturra e *C. eugenoides* CCC68 e aquelas de *C. canephora* depositadas no Coffee Hub e no NCBI foram avaliadas. Esse procedimento foi adotado para estimar quanto de variabilidade pode ter origem em artefatos de sequenciamento ou de montagem das sequências homólogas e homeólogas identificadas nos diferentes genomas. Um fragmento (aminoácidos 300 a 400, dos 596 que foram avaliados) de cada uma das 44 sequências de peptídeos deduzidos para domínios LRR, em alinhamento, pode ser visto na Figura 2.

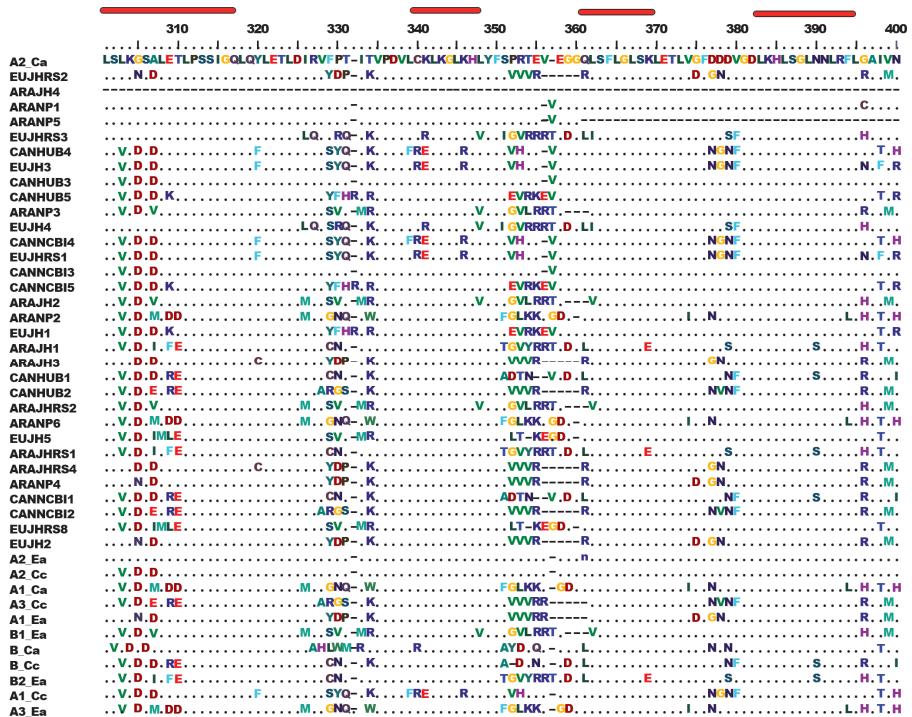


Figura 2. Fragmentos de domínios LRR em proteínas R do locus S_H3 de cafeeiros. As letras em diferentes cores são os códigos para os diferentes aminoácidos e os pontos são indicações de que o aminoácido encontrado na primeira sequência do alinhamento A2Ca repete-se nas demais. Nos nomes das sequências (extrema esquerda da figura), EU = *C. eugenioides*, ARA = *C. arabica*, CAN = *C. canephora* DH200-94. JH = sequências de *C. arabica* cv. Caturra e *C. eugenioides* depositadas no banco de genes do NCBI pela Universidade Johns Hopkins. RS = sequência identificada como *reference sequence* no NCBI. NP = sequência presente no genoma de *C. arabica* silvestre de origem etíope (*CaOE*). A_ e B_ = sequências publicadas por Ribas et al. (2011), para *C. arabica* cv. IAPAR59 (Ca e Ea) ou *C. canephora* IF200 (Cc). Motivos LRR, ricos no aminoácido leucina e importantes para o reconhecimento dos efetores Avr de *Hemileia vastatrix*, estão indicados sobre o alinhamento por traços em cor laranja. Os peptídeos ANANP5, ARAJH4 e B2-Ea e A2-Ea podem ser pseudogenes.

Um dos motivos LRR presente nos peptídeos deduzidos está representado na Figura 3A e a distribuição de motivos LRR na sequência do peptídeo CANHUB5, de *C. canephora* DH200-94 está representada na Figura 3B.

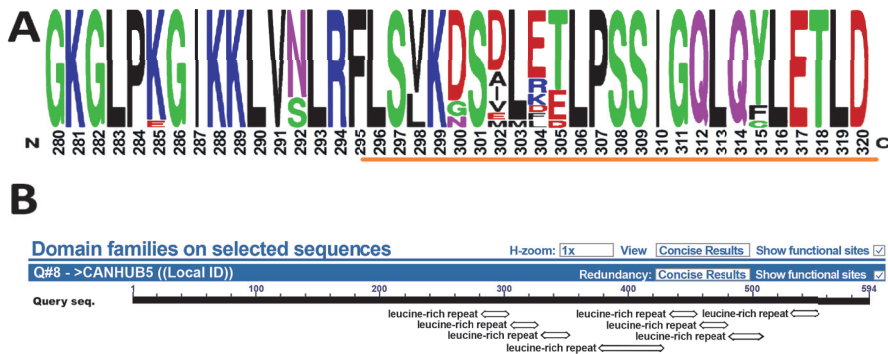


Figura 3. Sequência e organização de motivos LRR em proteínas CC-NBS-LRR reunidas em loci S_{H3} de cafeeiros. (A) A altura das letras que indicam os diferentes aminoácidos corresponde à frequência deles em cada posição (números abaixo da figura) no alinhamento das 44 sequências analisadas (vide Figura 2, primeiro motivo LRR identificado pelo traço laranja). L é o código de uma letra para o aminoácido leucina, que foi, nas posições 296, 306, 313, 316 e 319 preservado em todos os peptídeos analisados para este trabalho. A frequência alta de aminoácidos leucina é a característica primordial dos motivos LRR. (B) organização dos oito motivos LRR (setas) no peptídeo deduzido para o gene CANHUB5/CANNCBI5 codificado no cromossoma 3 de *C. canephora* DH200-94.

Na Figura 4, observa-se o agrupamento de 42 peptídeos (foram excluídos do agrupamento ARANP5 e ARAJH4, que parecem ser pseudogenes com códons de terminação muito precoces) representantes dos domínios LRR (regiões carboxiterminais das proteínas CC-NBS-LRR), segundo a semelhança/identidade entre os aminoácidos encontrados em cada posição (de 1 a 596). O agrupamento resultou na composição de seis ou sete subgrupos principais. Nos subgrupos mais populosos, como o 3 e o 6, peptídeos de *C. arabica* e *C. canephora* ficaram reunidos. Os subgrupos 2 e 5 são restritos a sequências codificadas nos genomas de *C. eugenioides* e *C. canephora*. Subgrupo espécie-exclusivo foi formado por peptídeos de *C. eugenioides* (subgrupo 4). Para *C. arabica*, o subgrupo 7 contou com contribuições de peptídeos dos genomas de *CaOE* e de *IAPAR59*. Ademais, sequências do S_{H3} de um mesmo genótipo ficaram dispersas pelos diversos subgrupos. Esse resultado pode indicar que os peptídeos ficaram agrupados principalmente de acordo com a especificidade/função, preservada durante a evolução ou a seleção de plantas resistentes em programas de melhoramento e não de acordo com o genótipo ou espécie em que foram encontrados. O agrupamento de acordo com o genótipo indicaria a ocorrência mais significativa de diversificação paralela e independente em cada espécie ou genótipo.

O subgrupo 2 (Figura 4) reuniu todas as sequências codificadas nos cromossomas 6 de *C. canephora* e *C. eugenioides*. A exceção nesse subgrupo foi a presença de um peptídeo de *C. eugenioides* (EUJH1, não referência) que está codificado no cromossoma 3. No entanto, a sequência de referência correspondente – EUJHRS1 – ficou incluída no subgrupo 5. Essa conformação sugere que, na sequência EUJH1, existe alguma falha de sequenciamento ou montagem que foi corrigida para gerar a sequência de referência EUJHRS1, pois a cópia do gene *R* denominada *EU..1* está localizada no cromossoma 3 como as demais sequências que codificam peptídeos do subgrupo 5, e não no cromossoma 6 como as sequências que codificam peptídeos do subgrupo 2.

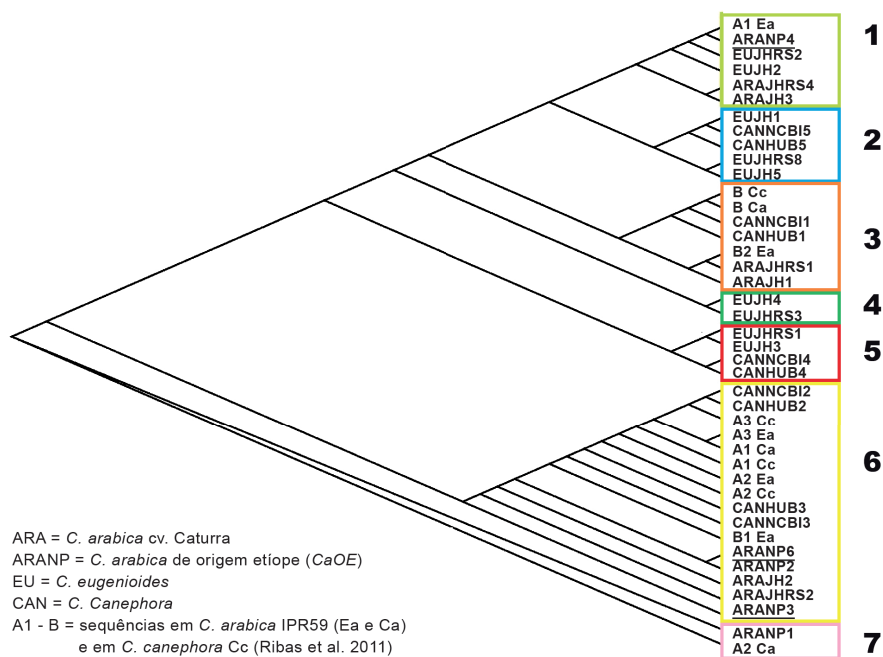


Figura 4. Agrupamento de regiões carboxiterminais de peptídeos deduzidos a partir da sequência de nucleotídeos que codificam proteínas R da família CC-NBS-LRRs agrupadas no locus S_H3 de cafeeiros. O agrupamento foi realizado utilizando o algoritmo de Fitch-Margoliash para reunir as sequências de acordo com a semelhança entre os aminoácidos que as constituem, sem considerar mecanismos evolutivos que interferem na substituição de um aminoácido por outro em posições correspondentes de cada sequência. Sequências ARANP (sublinhadas) são codificadas em cromossomas oriundos de *C. eugenioides* em *C. arabica* silvestre de origem etíope (CaOE). Sequências ARANP (não sublinhadas) são codificadas em cromossomas oriundos de *C. canephora*.

Sequências de *C. arabica* cv. IAPAR59 publicadas por Ribas et al. (2011) e de *CaOE*, ficaram distribuídas por quatro (1, 3, 6 e 7) e três (1, 6 e 7) subgrupos, respectivamente (Figura 4). No subgrupo 1, ficaram as sequências ARANP4 e A1-Ea codificadas no complemento cromossômico A^E e, no subgrupo 7, as sequências ARANP1 e A2-Ca codificadas no complemento A^C, tanto em IAPAR59 quanto em *CaOE*. Segundo Rodrigues Junior (1975) e Bettencourt e Rodrigues (1988) citados por Silva et al. (2006), cafeeiros arabica puros de origem etíope podem ter alelos S_H1, S_H2, S_H4 e S_H5, mas o S_H3 não é mencionado. Os peptídeos deduzidos para os genes de *C. arabica* cv. Caturra também se distribuíram pelos subgrupos 1, 3 e 6. O número de proteínas *R* codificadas no locus S_H3 da cultivar Caturra e naquele de *CaOE* é menor do que o encontrado no S_H3 da cultivar IAPAR59. Na cultivar Caturra, classificada como hiper suscetível à ferrugem (Zambolim, 2016), as quatro cópias de genes *R* do locus S_H3 estão no cromossoma 3 de ancestral *C. canephora* (A^C) e uma dessas cópias – *ARAJH4* – aparentemente não é funcional, devido à ocorrência de um códon precoce de terminação da tradução. No subgrupo 1, também estão sequências EUJH2/EUJHRS2 codificadas no cromossoma 3 de *C. eugenoides* cuja composição de aminoácidos é muito similar nos depósitos de referência e não referência disponíveis no NCBI.

Interessante notar que os quatro peptídeos LRR codificados nos cromossomas 3 de *C. canephora* DH200-94 (CAN..1-3) e de IF200 (A1 até A3-Cc e B-Cc de Ribas et al., 2011) ocuparam principalmente os subgrupos 3 e 6 (Figura 4). Os genes S_H6 a S_H9 foram identificados em *C. canephora* (Rodrigues Junior, 1975; Bettencourt; Rodrigues, 1988, citados por Zambolim, 2016). A exceção a esse arranjo foi CAN..4, que ficou no subgrupo 5, no qual ocorreu uma divergência de sequências segundo os genótipos *C. canephora* ou *C. eugenoides*. Duas das sequências em cada uma dessas espécies são mais semelhantes entre si do que as restantes da mesma espécie e foram reunidas em um mesmo grupo: EUJH1 codificada no cromossoma 3 (mas vide a menção anterior sobre a discrepância entre EUJH1 e EUJHRS1, neste mesmo trabalho) e EUJH5/EUJHRS8 codificadas no cromossoma 6 não referência e referência de *C. eugenoides*, respectivamente, que ficaram no subgrupo 2 e CANNCB12/CANHUB2 e CANNCB13/CANHUB3, codificadas no cromossoma 3 de *C. canephora* DH200-94, que ficaram no subgrupo 6.

Em *C. canephora* DH200-94, CANHUB1 é a cópia mais próxima da extremidade do cromossomo 3 e é extremamente similar a CANNCB11,

que é codificada no *contig* com numeração de *scaffold* mais baixa no NCBI. CANHUB1 e CANNB1 são extremamente parecidas e ficaram no subgrupo 3 (Figura 4) e assim por diante com os peptídeos deduzidos para todos os genes no S_H3 desse genótipo, que apresentaram-se quase idênticos nos dois bancos de dados, Coffee Hub e NCBI. Essa correspondência perfeita denota a excelente qualidade do sequenciamento e processamento dos dados referentes ao genoma de *C. canephora* DH200-94. Essa qualidade permite considerar que os pares de peptídeos deduzidos CANNB2/CANHUB2 e CANNB3/CANHUB3 apresentem especificidade semelhante quanto à interação com efetores de patógenos porque foram os dois únicos pares de sequências de *C. canephora* DH200-94 que ocuparam juntos um mesmo subgrupo (o de número 6).

Para sequências de *C. eugenioides* e *C. arabica* var. Caturra, entretanto, serão necessárias mais análises antes que hipótese semelhante possa ser considerada. As sequências referência (JHRS) na Figura 4 e não referência (JH apenas) de *C. eugenioides* nem sempre foram agrupadas em um mesmo subgrupo, já que há muitas divergências entre os diferentes depósitos nos bancos de genes. Como exemplos, EUJH1 que ficou no subgrupo 2 enquanto a EUJHRS1 está no subgrupo 5 como já mencionado, e EUJH3 no subgrupo 5, enquanto EUJHRS3 está no subgrupo 4. EUJH5 e EUJHRS8 são ambas codificadas pelas cópias “mais a 3” no cromossoma 6 de *C. eugenioides*, mas no cromossoma de referência depositado no NCBI há, pelo menos, três fragmentos de genes com homologia quase completa a partes do domínio LRR que estão ausentes do cromossoma 6 não referência do mesmo genótipo.

Sabendo-se que as combinações das variantes estruturais dos genes incluídos no locus S_H3 apresentam potencial para explicar parte das reações de defesa das diferentes variedades de cafeeiros contra diferentes raças de *H. vastatrix*, espera-se, em experimentos futuros com a inclusão de cafeeiros S_H3^+ , como *C. liberica* (Prakash et al., 2004; Zambolim, 2016), e análise do padrão de expressão dos diferentes genes dos agrupamentos quando desafiados pelos fungos, contribuir para a compreensão de como isso ocorre.

Conclusões

A conformação dos agrupamentos de genes R que codificam proteínas da família CC-NBS-LLR, encontrados nos loci S_H3 de seis genótipos de cafeeiros, permite afirmar que os loci são complexos e multialélicos no gênero *Coffea*. Parte da diversidade nesses loci é consequência de artefatos de sequenciamento e montagem dos dados dos genomas, e parte é real e deve estar relacionada, em conjunto com outros fatores, com a complexidade das interações entre os cafeeiros e as diversas raças de *H. vastatrix*.

Referências

- ALKIMIM, E. R.; CAIXETA, E. T.; SOUSA, T. V.; PEREIRA, A. A.; OLIVEIRA, A. C. B. de; ZAMBOLIM, L.; SAKIYAMA, N. S. Marker-assisted selection provides arabica coffee with genes from other *Coffea* species targeting on multiple resistance to rust and coffee berry disease. **Molecular Breeding**, v. 37, article number 6, 2017. DOI 10.1007/s11032-016-0609-1.
- ANGELO, Paula Cristina S. (2021), "CC-NBS-LRR GENE SEQUENCES", **Mendeley Data**, V1, doi: 10.17632/sp2zhxd4jj.1
- BARKA, G. D.; CAIXETA, E. T.; ALMEIDA, R. F. de; ALVARENGA, S. M.; ZAMBOLIM, L. Differential expression of molecular rust resistance components have distinctive profiles in *Coffea arabica* – *Hemileia vastatrix* interactions. **European Journal of Plant Pathology**, v. 149, n. 3, p. 543-561, 2017. DOI 10.1007/s10658-017-1202-0.
- BONARDI, V.; DANGL, J. L. How complex are intracellular immune receptor signaling complexes? **Frontiers in Plant Science**, v. 3, 2012. DOI 10.3389/fpls.2012.00237.
- CABRAL, P. G. C.; ZAMBOLIM, E. M.; ZAMBOLIM, L.; LELIS, T. P.; CAPUCHO, A. S.; CAIXETA, E. T. Identification of a new race of *Hemileia vastatrix* in Brazil. **Australasian Plant Disease Notes**, v. 4, p. 129-130, 2009. DOI 10.1071/DN09052.
- CAPUCHO, A. S.; ZAMBOLIM, E. M.; FREITAS, R. L.; HADDAD, F.; CAIXETA, E. T.; ZAMBOLIM, L. Identification of race XXXIII of *Hemileia vastatrix* on *Coffea arabica* Catimor derivatives in Brazil. **Australasian Plant Disease Notes**, v. 7, p. 189-191, 2012. DOI 10.1007/s13314-012-0081-7.
- CARVALHO, A.; MONACO, L. C.; FAZUOLI, L. C.; RIBEIRO, I. J. A. Transferência de fatores genéticos de resistência a *Hemileia vastatrix* para o cultivar Mundo Novo. **Bragantia**, v. 36, n. 1, p. 93-102, 1977.
- CARVALHO, C. R.; FERNANDES, R. C.; CARVALHO, G. M. A.; BARRETO, R. W.; EVANS, H. C. Cryptosexuality and the genetic diversity paradox in Coffee Rust, *Hemileia vastatrix*. **PLoS ONE**, v. 6, n. 11, e26387, 2011. DOI 10.1371/journal.pone.0026387.
- CONCEIÇÃO, A. S. da.; FAZUOLI, L. C.; BRAGHINI, M. T. Avaliação e seleção de progênies F3 de cafeeiros de porte baixo com o gene S_H3 de resistência a *Hemileia vastatrix* Berk. et Br. **Bragantia**, v. 64, n. 4, p. 547-559, 2005. DOI 10.1590/S0006-87052005000400004.
- FLOR, H. H. Current status of the gene-for-gene concept. **Annual Review of Phytopathology**, v. 9, p. 275-296, 1971. DOI 10.1146/annurev.py.09.090171.001423.

- HALL, T. A. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. **Nucleic Acids Symposium Series**, n. 41, p. 95-98, 1999.
- MAHÉ, L.; COMBES, M.-C.; VÁRZEA, V. M. P.; GUILHAUMON, C.; LASHERMES, P. Development of sequence characterized DNA markers linked to leaf rust (*Hemileia vastatrix*) resistance in coffee (*Coffea arabica* L.). **Molecular Breeding**, v. 21, n. 1, p. 105-113, 2008. DOI 10.1007/s11032-007-9112-z.
- MATIELLO, J. B.; SANTINATO, R.; ALMEIDA, S. R.; GARCIA, A. W. R. **Cultura de café no Brasil: manual de recomendações**. São Paulo: Futurama, 2016. 584 p.
- NEPAL, M. P.; ANDERSEN, E. J.; NEUPANE, S.; BENSON, B. V. Comparative genomics of Non-TNL disease resistance genes from six plant species. **Genes**, v. 8, n. 10, 2017. DOI: 10.3390/genes8100249.
- PRAKASH, N. S.; MARQUES, D. V.; VARZEA, V. M. P.; SILVA, M. C.; COMBES, M. C.; LASHERMES, P. Introgression molecular analysis of a leaf rust resistance gene from *Coffea liberica* into *C. arabica* L. **Theoretical and Applied Genetics**, v. 109, n. 6, p. 1311-1317, 2004. DOI: 10.1007/s00122-004-1748-z.
- RIBAS, A. F.; CENCI, A.; COMBES, M.-C.; ETIENNE, H.; LASHERMES, P. Organization and molecular evolution of a disease-resistance gene cluster in coffee trees. **BMC Genomics**, v. 12, p. 1-12, 2011. DOI: 10.1186/1471-2164-12-240.
- SCHIEBER, E. Economic impact of coffee rust in Latin America. **Annual Review of Phytopathology**, v. 10, p. 491-510, 1972. DOI: 10.1146/annurev.py.10.090172.002423.
- SERA, G. H.; SERA, T.; ITO, D. S.; AZEVEDO, J. A. de; MATA, J. S. da; DOI, D. S.; RIBEIRO FILHO, C. Selection for durable resistance to leaf rust using testcrosses on IAPAR-59 and Tupi IAC 1669-33 cultivars of *Coffea arabica*. **Brazilian Archives of Biology and Technology**, v. 50, n. 4, p. 565-570, 2007.
- SERA, G. H.; SERA, T.; ITO, D. S.; FONSECA, I. C. de B.; KANAYAMA, F. S.; GROSSI, L. Del; SHIGUEOKA, L. H. Seleção para a resistência à ferrugem em progênies das cultivares de café IPR 99 e IPR 107. **Bragantia**, v. 69, n. 3, p. 547-554, 2010.
- SILVA, M. do C.; VÁRZEA, V.; GUERRA-GUIMARÃES, L.; AZINHEIRA, H. G.; FERNANDEZ, D.; PETITOT, A.-S.; BERTRAND, B.; LASHERMES, P.; NICOLE, M. Coffee resistance to the main diseases: leaf rust and coffee berry disease. **Brazilian Journal of Plant Physiology**, v. 18, n. 1, p.119-147, 2006. DOI 10.1590/S1677-04202006000100010.
- TALHINHAS, P.; BATISTA, D.; DINIZ, I.; VIEIRA, A.; SILVA, D. N.; LOUREIRO, A.; TAVARES, S.; PEREIRA, A. P.; AZINHEIRA, H. G.; GUERRA-GUIMARÃES, L.; VARZEA, V.; SILVA, M. do C. The coffee leaf rust pathogen *Hemileia vastatrix*: one and a half centuries around the tropics. **Molecular Plant Pathology**, v. 18, n. 18, p.1039-1051, 2017. DOI 10.1111/mpp.12512.
- ZAMBOLIM, L. Current status and management of coffee leaf rust in Brazil. **Tropical Plant Pathology**, v. 41, p. 1-8, 2016. DOI 10.1007/s40858-016-0065-9.

Embrapa

Café



MINISTÉRIO DA
AGRICULTURA, PECUÁRIA
E ABASTECIMENTO

