

VARIABILIDADE NO DOMÍNIO LRR DE PEPTÍDEOS CODIFICADOS EM “CLUSTERS” HOMÓLOGOS AO S_{H3} EM DIFERENTES GENOMAS DE *Coffea*¹

Paula Cristina da Silva Angelo²; Caroline Ariyoshi³; Eveline Teixeira Caixeta⁴; Luiz Filipe Protásio Pereira⁵; Alessandra F. Ribas⁶; Gustavo H. Sera⁷

¹ Trabalho financiado pelo Consórcio Brasileiro de Pesquisa e Desenvolvimento do Café – Consórcio Pesquisa Café

² Pesquisadora, D.Sc., Embrapa Café/Instituto Agronômico do Paraná - IAPAR, Londrina-PR, paula.angelo@embrapa.br

³ Doutoranda, Universidade Estadual de Londrina/IAPAR, Londrina-PR, carolinearyoshi@hotmail.com

⁴ Pesquisadora, Ph.D., Embrapa Café/Universidade Federal de Viçosa, Viçosa-MG, eveline.caixeta@embrapa.br

⁵ Pesquisador, Ph.D., Embrapa Café/IAPAR, Londrina-PR, filipe.pereira@embrapa.br

⁶ Professora, Ph.D., Universidade do Oeste Paulista, Presidente Prudente-SP, alessandra_ribas@hotmail.com

⁷ Pesquisador, Ph.D., IAPAR, Londrina-PR, gustavosera@iapar.br

RESUMO: O locus S_{H3} é relacionado com as reações de defesa contra *Hemileia vastatrix*, o agente causador da ferrugem alaranjada. Há um “cluster” de genes *CC-NBS-LRR* no locus S_{H3} da variedade IAPAR59 de *Coffea arabica*. Esse “cluster” reúne oito cópias de *CC-NBS-LRRs*, que variam na sua estrutura primária, especialmente na região carboxi-terminal onde são localizados motivos ricos em leucina. Esses motivos LRR são característicos de proteínas que participam do reconhecimento de patógenos pelas plantas. Os “clusters” homólogos na variedade Caturra e em um cafeeiro de origem etíope da espécie *C. arabica*, em *C. canephora* DH200-94 e em *C. eugenioides* CCC68 foram identificados. A variabilidade das regiões codificadoras incluídas no “cluster” S_{H3} foi avaliada alinhando com o aplicativo CLUSTAW e agrupando pelo método de Fitch-Margoliash as sequências da região carboxi-terminal dos peptídeos deduzidos de todas as cópias de *NBS-LRRs* codificadas em “clusters” homólogos naqueles quatro genótipos e também previamente analisadas em IAPAR59 e em *C. canephora* IF200. Concluiu-se que há divergência entre as sequências dos domínios LRR de diferentes membros dos “clusters” em diferentes genótipos, o que pode ser parcialmente resultante dos processos de sequenciamento e montagem, mas que também resulta de evolução e seleção. Também há diferença no número de cópias do gene em diferentes genótipos. Consideramos a hipótese de que o número, a organização, a integridade e a estrutura primária dos genes nos “clusters” S_{H3} podem estar relacionados com o mecanismo de reconhecimento das raças fisiológicas de *H. vastatrix* por diferentes variedades de cafeeiros.

PALAVRAS-CHAVE: doença do cafeeiro, polimorfismo estrutural, isoformas protéicas.

LRR DOMAIN VARIABILITY FOR PEPTIDES ENCODED IN GENE CLUSTERS HOMOLOGOUS TO THE S_{H3} IN DIFFERENT *Coffea* GENOMES.

ABSTRACT: The S_{H3} locus is involved in the resistance to *H. vastatrix*, which causes the leaf rust disease. There is a cluster of *CC-NBS-LRR* genes which are encoded in the S_{H3} locus of *Coffea arabica* cv. IAPAR59. In IAPAR59 the cluster is composed by eight copies of the gene, which display variations in the primary structure, specially at the carboxy-terminal region, where leucine rich repeats are concentrated. These LRR motives are characteristic of proteins enrolled in the plant-pathogen recognition mechanisms. The homologous clusters were identified in recently sequenced *C. arabica* cv. Caturra and in a coffee plant native to Ethiopia, in *C. canephora* DH200-94 and in *C. eugenioides* CCC68 genomes. The variability in the carboxy-terminal region of the polypeptides deduced from the genes encoded in the S_{H3} clusters of these four genotypes and those previously analyzed that are encoded in the IAPAR59 and in the *C. canephora* IF200 homologous clusters was evaluated by aligning the amino acid sequences with CLUSTAW and grouping them by the Fitch-Margoliash method. It was concluded that some of the diversity observed in the LRR domains of different cluster members is due to the processes of sequencing and assembling the genomes, but there is diversity resultant from evolution and selection as well. There are differences regarding the gene copy numbers in different genotypes as well. We are considering the hypothesis that the number, the organization, the integrity and the primary structure of the genes in the S_{H3} clusters can be related to the recognition of *H. vastatrix* physiological races by the different coffee plant varieties.

KEY WORDS: coffee disease, structural polymorphism, protein isoforms.

INTRODUÇÃO

A ferrugem alaranjada é uma das principais doenças do cafeeiro arábica, reduzindo produtividade devido a desfolha das plantas e morte de ramos (Matiello et al. 2016; Zambolim 2016). O agente patogênico causador da doença é o fungo *Hemileia vastatrix* e o conhecimento das reações de interação desse fungo com os cafeeiros tem avançado constantemente (Barka et al., 2017).

O locus SH3 da cultivar IAPAR59 foi clonado e sequenciado por Ribas et al. (2011). É uma região que tem cerca de 250 kb onde um “cluster” com oito genes de resistência foi localizado, estando cinco deles no complemento oriundo de *C. eugenoides* e três no complemento oriundo de *C. canephora*. No genoma de *C. canephora* o cluster homólogo é composto por quatro genes de resistência. Cada um dos genes de resistência que compõem o “cluster” codifica uma proteína do tipo CC-NBS-LRR (Ribas et al., 2011).

As proteínas dessa família são reconhecidas participantes das reações de defesas do tipo “gene a gene” ou, alternativamente, do tipo “gene de guarda”, em diversas espécies vegetais. No primeiro modelo de reação de defesa, há interação direta entre a proteína R de defesa da planta e um efetor *Avr* específico de cada raça do patógeno. No segundo modelo, a proteína de defesa interage com outras proteínas da planta que são expressas como reação à infecção pelo fungo. As CC-NBS-LRRs são assim denominadas porque apresentam um domínio “coiled-coil” (CC), um domínio central de ligação a nucleotídeos (“nucleotide binding sequence” NBS) e um domínio carboxi-terminal que contém motivos ricos em leucina, denominados LRR para “leucine rich repeats”, em número variável. Esse último domínio é o que apresenta maior variabilidade de estrutura primária e é também o domínio que interage e reconhece os 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 CC-NBS-LRRs é atribuído à necessidade de interagir com patógenos que têm ciclo de vida reduzido e grande capacidade de produzir *Avrs* alternativos (Nepal et al., 2017).

Para o presente trabalho, “clusters” homólogos ao SH3 identificado por Ribas et al. (2011), que contêm as regiões codificadoras das CC-NBS-LRR incluindo o pequeno intron presente em cada uma das cópias, foram identificados nos genomas de *C. arabica*, *C. canephora* e *C. eugenoides* sequenciados recentemente. As combinações das variantes estruturais dos genes incluídos especificamente nesse “cluster” apresentam potencial para explicar parte das reações de defesa das diferentes variedades de *C. arabica* contra diferentes raças de *H. vastatrix*. O objetivo desse trabalho foi avaliar a diversidade da estrutura primária do domínio LRR das CC-NBS-LRR codificadas nos genomas de *Coffea* sequenciados, em “clusters” homólogos ao que foi denominado SH3 em IAPAR59.

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. eugenoides*. A partir das sequências de DNA genômico recuperadas, foram deduzidos os peptídeos codificados pela região carboxi-terminal (596 resíduos de aminoácidos da região carboxi-terminal, incluindo o domínio LRR) de cada cópia dos genes codificadores de CC-NBS-LRRs encontrada nos “clusters” SH3 dos diferentes genomas. Essas regiões foram alinhadas e agrupadas utilizando o algoritmo de Fitch-Margoliash (taxa de substituição livre ao longo de toda a sequência; version 3.5c, Phylogeny Inference Package – PHYLIP, <http://evolution.gs.washington.edu/phylip.html>) incluído no aplicativo BIOEDIT v.7.2.6 (Hall, 1999)]. Os bioprojetos consultados no NCBI (www.ncbi.nlm.nih.gov) foram *C. arabica* cv. Caturra vermelho, PRJNA 497895 sequências referência e não referência; *C. eugenoides* acesso CCC68 do Cenicafé, Colômbia, PRJNA 508372 e PRJNA 497891 sequências referência e não referência; *C. canephora* diaploide DH200-94 IRD França, PRJEB 4211. As sequências desse mesmo genótipo de *C. canephora* depositadas no Coffee Hub (www.coffeehub.org) também foram acessadas. E por fim, foram acessadas sequências do genoma de um indivíduo de *C. arabica*, oriundo de uma população natural da Etiópia, mantido *ex situ* no Instituto Agrônomo do Paraná – IAPAR – Londrina (*CaOE*, sequências não publicadas). A codificação dessas sequências no presente trabalho é: **ARA** para sequências de *C. arabica*, depositadas no NCBI pela John Hopkins University, classificadas como “reference sequences” (RF) ou não. **ARANP** para *CaOE*, *C. arabica* de origem etíope. **CAN** para sequências de *C. canephora*, depositadas no NCBI ou Coffee Hub (HUB). **EU** para sequências de *C. eugenoides*, depositadas no NCBI pela John Hopkins University, classificadas como “reference sequences” (RF) ou não.

RESULTADOS E DISCUSSÃO

Em todos os genomas acessados, a região homóloga ao “clusters” SH3 de *C. arabica* cv. IAPAR59 sequenciada por Ribas et al. (2011) foi localizada nos cromossomos de número 3. Diferentes números de cópias dos genes codificando proteínas CC-NBS-LRR foram encontrados em “clusters” homólogos nos diferentes genomas.

Em *C. arabica* de origem etíope (*CaOE*) foi observada, o “cluster” apresenta organização semelhante à registrada para a cultivar IAPAR59, com ausência de uma das três cópias do gene codificador de CC-NBS-LRR no complemento oriundo de *C. canephora* e de uma das cinco cópias no complemento *C. eugenoides* (Fig. 1). Portanto, o “cluster” em *CaOE* é formado por duas regiões codificadoras de CC-NBS-LRR no complemento oriundo de *C. canephora* (A^C) e quatro regiões codificadoras no complemento oriundo de *C. eugenoides* (A^E), todas nos cromossomos 3. Observou-se preservação da orientação das cópias do gene, o número e o tamanho dos espaçadores de 27,3 kb em A^C e 168,4, 14,8 e 26,9 kb em A^E , muito semelhantes em *CaOE* e na cv. IAPAR59. A estrutura de cada cópia do gene inserida no “cluster” também é bastante preservada, com o exon 1 variando de 1040 a 1045 pb, introns em torno de 250 pb e o exon 2 apresentando entre 1773 e 1795 pb. Uma das cópias do gene, ARANP5 (Fig. 1, seta 5) no complemento A^E em *CaOE*, tem, aparentemente, um códon de terminação precoce e pode ser um pseudogene.

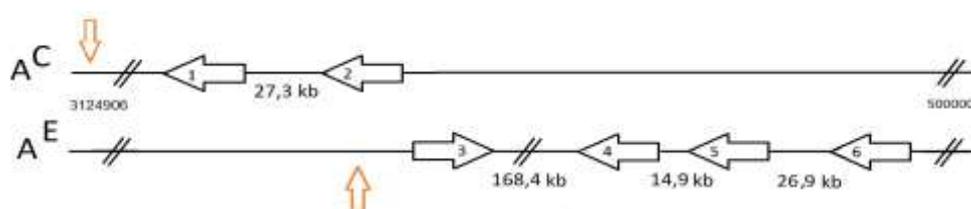


Figura 1. Organização do “cluster” de genes de resistência homólogo ao S_H3 em *C. arabica* de origem etíope (*CaOE*). As setas pretas indicam a orientação e a localização de cada uma das seis cópias de regiões codificadoras de CC-NBS-LRR que compõem o “cluster”. As setas vermelhas indicam a posição das cópias do gene que foram observados na cultivar IAPAR59 mas estão ausentes do genoma sequenciado e ainda não publicado de *CaOE*. Os números nas extremidades da figura indicam pares de base. A^C = cromossomo 3 do complemento oriundo de *C. canephora*. A^E = cromossomo 3 do complemento oriundo de *C. eugenoides*. A representação não está em escala e portanto as posições são aproximadas.

No cromossomo 3 de *C. eugenoides*, foram encontradas quatro cópias do gene e espaçadores de 219, 15 e 16 kb, aproximadamente. Em *C. canephora*, foram encontradas quatro cópias no cromossomo 3, espaçadas por 226, 16 e 11 kb. Em cada um desses dois últimos genótipos, uma cópia isolada do gene foi localizada nos cromossomos 6, fora do “cluster” onde as demais cópias se organizaram. Essas cópias “extras” precisam ser analisadas com mais cuidado para averiguar se são cópias expressas ou pseudogenes sem promotores. 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. O número de cópias do gene no “cluster” da cultivar Caturra e naquele de *CaOE* é menor do que o número de cópias na variedade IAPAR59. Segundo Bettencourt e Rodrigues (apud. Silva et al., 2006) cafeeiros *arabica* puros de origem etíope podem ter alelos S_H1 , S_H2 , S_H4 e S_H5 . A variedade Caturra de *C. arabica* tem o gene S_H5 de resistência à raça II do fungo (Bettencourt e Noronha-Wagner, 1971). Em *C. canephora* var. conilon foram identificadas três raças fisiológicas do fungo, sendo I(v2v5), II(v5) and III(v1v5) (Zambolin, 2016).

A diversidade da estrutura primária foi avaliada para 42 sequências de regiões carboxi-terminais (596 aminoácidos em peptídeos deduzidos da sequência dos genes) de CC-NBS-LRRs. Nessa análise foram incluídos peptídeos deduzidos para todas as cópias citadas acima, codificadas nos cromossomos 3 e 6 de *C. arabica*, *C. canephora* e *C. eugenoides* disponibilizadas no NCBI e ainda não publicadas. Também foram incluídas na análise as oito cópias identificadas no genoma de IAPAR59 (contando com A2-E^a e B2-E^a, que apresentam códons de terminação precoces) e as quatro cópias 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* e aquelas de *C. canephora* depositadas no Coffee Hub e no NCBI foram incluídas na análise. Esse procedimento foi adotado para estimar o quanto de variabilidade pode ter origem em artefatos de sequenciamento ou de montagem das sequências homólogas e homeólogas identificadas nos diferentes genomas.

Uma parte do alinhamento entre as 44 sequências pode ser visto na Fig. 2. A região carboxi-terminal dos peptídeos deduzidos foi selecionada por sua importância para a especificidade da interação dos cafeeiros com as diferentes raças de *H. vastatrix*. Essa região das proteínas NBS-LRR tem sido associada com a especificidade da interação com os fatores *Avr* dos patógenos. Parte da especificidade planta-patógeno deve-se certamente à variabilidade da sequência e à organização de motivos LRR nos diferentes membros dessa família de proteínas. Um dos motivos LRR presente em peptídeos deduzidos codificados no “cluster” S_H3 está representado na Fig. 3a e a distribuição de motivos LRR na sequência CANHUB5 está representada na Fig. 3b.



Figura 2. Fragmento do domínio LRR em CC-NBS-LRR de cafeeiro. EU = *C. eugenioides*, ARA = *C. arabica*, CAN = *C. canephora*. A1-A3, B e Cc = seqüências publicadas por Ribas et al. (2011), para *C. arabica* cv. IPR59 e *C. canephora* Cc. Motivos LRR estão indicados sobre o alinhamento com traços em cor laranja. ANANP5, ARAJH4 e B2-Ea e A2-Ea podem ser pseudogenes.



Figura 3. Seqüência e organização de motivos LRR em CC-NBS-LRR reunidos no “cluster” homólogo da S_H3 . **A** – A altura das letras que indicam os diferentes aminoácidos na seqüência correspondem à frequência deles em cada posição do alinhamento das 44 seqüências analisadas (vide Fig. 2). L é o código de uma letra para a leucina. **B** – organização das repetições LRR no peptídeo deduzido para as seqüências CANHUB5/CANNCB15 codificado no cromossomo 3 de *C. canephora*.

O agrupamento das seqüências (Fig. 4), excluídos ANANP5 e ARAJH4 que parecem ser pseudogenes com codons de terminação muito precoces, resultou na composição de seis ou sete clados principais. Nos clados mais amplos, como o 3 e o 6, peptídeos de *C. arabica* e *C. canephora* estão reunidos. Os clados 2 e 5 são restritos a seqüências codificadas nos genomas de *C. eugenioides* e *C. canephora*. E clados espécie-exclusivos foram formados por peptídeos de *C. eugenioides* (clado 4) e de *C. arabica*, com contribuição de *CaOE* e de *IAPAR59* (clado 7).

O clado 2 reuniu todas as sequências codificadas nos cromossomos 6 de *C. canephora* e *C. eugenioides*. A exceção nesse clado foi a presença de um peptídeo de *C. eugenioides* (EJH1, não referência) que está codificado no cromossomo 3.

Sequências de *C. arabica* cv. IAPAR59 e de *CaOE* ficaram distribuídas por quatro (1, 3, 6 e 7) e três clados (1, 6 e 7), respectivamente. No clado 1, ficaram as sequências ARANP4 e A1-Ea codificadas no complemento A^E e, no clado 7, as sequências ARANP1 e A2-Ca codificadas no complemento A^C, tanto em IAPAR59 quanto em *CaOE* (Fig. 4). No clado 1 também estão sequências EUJH2/EUJHRS2 codificadas no cromossomo 3 de *C. eugenioides* cuja composição de aminoácidos é muito similar nos depósitos de referência e não referência disponíveis no NCBI.

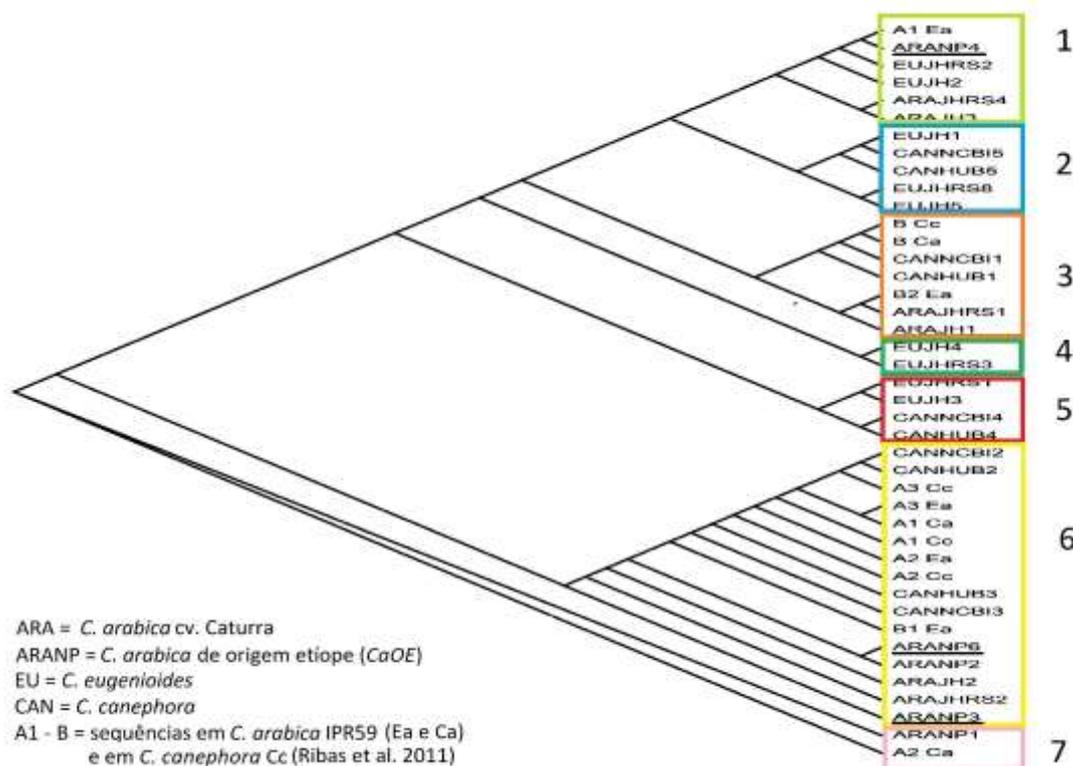


Figura 4. Agrupamento de regiões carboxi-terminais de peptídeos deduzidos a partir das sequências codificadoras de CC-NBS-LRRs de cafeeiros utilizando o algoritmo de Fitch-Margoliash. Sequências ARANP sublinhadas são codificadas no complemento ou subgenoma oriundo de *C. eugenioides* em *C. arabica* de origem etíope (*CaOE*), cujo genoma foi sequenciado mas ainda não publicado.

Interessante notar que as sequências codificadas em *C. canephora* distribuíram-se por quatro clados e o mesmo ocorreu com as sequências de *C. eugenioides*. Duas das sequências em cada uma dessas espécies são mais semelhantes uma à outra do que às restantes da mesma espécie: EUJH1 codificada no cromossomo 3, e EUJH5/EUJHRS8 codificadas no cromossomo 6 não-referência e referência de *C. eugenioides*, respectivamente, que ficaram agrupadas no clado 2 e CANNCB12/CANHUB2 e CANNCB13/CANHUB3, codificadas no cromossomo 3 de *C. canephora* que ficaram agrupadas no clado 6 (Fig. 4). Em *C. canephora*, cópias do gene que ocuparam posição similar no “cluster” com relação à posição/número das bases nos cromossomos sempre apresentaram estrutura idêntica ou muito similar nos diferentes acessos/depósitos de genomas. Por exemplo, CANHUB1 é a cópia codificada nas bases mais próximas da extremidade do cromossomo 3 e é extremamente similar a CANNCB11, que é codificada no “contig” de numeração mais baixa do scaffold entre aqueles que estão relacionados ao “cluster” de *C. canephora*. CANHUB1 e CANNCB11 são extremamente parecidas e ficaram agrupadas no clado 3 (Fig. 4) e assim por diante.

Por outro lado, as sequências “referência” e “não referência” de *C. eugenoides* nem sempre foram agrupadas em um mesmo clado porque há muitas divergências entre os diferentes depósitos. Exemplos são as sequências EUJH1 que ficou no clado 2 enquanto a EUJHRS1 está no clado 5, e EUJH3 no clado 5, enquanto EUJHRS3 está no clado 4. EUJH5 e EUJHRS8 são ambas codificadas pelas cópias “ mais a 3' ” no cromossomo 6 de *C. eugenoides*, mas no cromossomo de referência há pelo menos três fragmentos com homologia quase completa a partes do domínio LRR que, no entanto, estão ausentes do cromossomo 6 “não referência” do mesmo genótipo. Portanto, é possível considerar a hipótese de que os pares de peptídeos CANNB12/CANHUB2 e CANNB13/CANHUB3 sejam codificados pela cópia número 2 e pela cópia número 3 do gene, respectivamente, no “cluster” de *C. canephora* DH200-94 e que apresentem a mesma especificidade quanto à interação com efetores de patógenos porque ocuparam, os dois pares de sequências, o mesmo clado 6, a despeito dos membros de cada par serem mais semelhantes entre si do que semelhantes a membros de outros pares. Por outro lado, para sequências de *C. eugenoides* e *C. arabica* var. Caturra, mais análises serão necessárias antes que hipótese semelhante possa ser considerada.

CONCLUSÕES

Conclui-se que parte da diversidade entre CC-NBS-LRRs codificadas nos “clusters” homólogos a S_H3 pode ser devida à artefatos de sequenciamento ou montagem, mas, ainda assim, a variabilidade da estrutura primária e na organização/localização de motivos LRR e das cópias do gene nos “clusters” parece suficiente para explicar, pelo menos em parte e em conjunto com outros fatores, as interações entre o cafeeiro e *H. vastatrix*. Essa hipótese será investigada em ações de pesquisa que estão em andamento.

REFERÊNCIAS BIBLIOGRÁFICAS

- HALL, T.A. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids. Symp. Ser.*, v. 41, pp. 5-98, 1999.
- BARKA, G.D., CAIXETA, E.T., ALMEIDA, R.F. De CRUZ, C.D., ZAMBOLIM, L., PEREIRA, A.A. & OLIVEIRA, A.C.B. 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. 31, 2017.
- BETTENCOURT, A. J. & NORONHA-WAGNER, M. GENETIC FACTORS CONDITIONING RESISTANCE OF *COFFEA ARABICA* L. TO *Hemileia vastatrix* Berk & Br. *Agronomia Lusitana*, v. 31, pp 285-292, 1971.
- MATIELLO, J.B., SANTINATO, R., ALMEIDA, S.R. & GARCIA, A.W.R. *Cultura de Café no Brasil: manual de recomendações*. Futurama, São Paulo, 584 p., 2016.
- 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. 249, pp. 2-17, 2017.
- 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, n. 240, pp. 2-12, 2011.
- SILVA, M. D. 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. *Braz. J. Plant Physiol.*, v. 18, pp. 119-147, 2006.
- ZAMBOLIM, L. Current status and management of coffee leaf rust in Brazil. *Tropical Plant Pathology*, v. 41, pp. 1-8, 2016.