

ESTUDO DE ASSOCIAÇÃO GENÔMICA AMPLA PARA INTERAÇÃO de *Coffea arabica* L. E *Pseudomonas syringae* pv. *garcae*¹

Caroline Ariyoshi²; Gustavo César Sant'ana³; Gustavo Hiroshi Sera⁴; Lucas Mateus Rivero Rodrigues⁵; Lívia Maria Nogueira⁶; Rafaelle Vicchia Ferreira⁷; Mariane Silva Felício⁸; Bruna Silvestre Rodrigues da Silva⁹; Fernanda Freitas de Oliveira¹⁰; Larissa Giroto¹¹; Suzana Tiemi Ivamoto¹²; Douglas Silva Domingues¹³; Luiz Filipe Protasio Pereira¹⁴

¹Trabalho financiado pelo Consórcio Brasileiro de Pesquisa e Desenvolvimento do Café – Consórcio Pesquisa Café e pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes)

²Bolsista Capes, MS, carolineariyoshi@hotmail.com

³Pesquisador, PhD, Tropical Melhoramento e Genética, Cambé-PR, gustavocsantana80@gmail.com

⁴Pesquisador, PhD, Instituto Agrônomo do Paraná, Londrina-PR, gustavosera@iapar.br

⁵Bolsista INCT, PhD, lucasmriverorodrigues@gmail.com

⁶Bolsista Capes, PhD, livia.nogueira@hotmail.com

⁷Bolsista Capes, PhD, rafaelle.ferreira@gmail.com

⁸Bolsista Capes, MS, marianesfelicio@gmail.com

⁹Bolsista Consórcio Pesquisa Café, PhD, brunasilvestrerodrigues@hotmail.com

¹⁰Bolsista Capes, MS, fernandafreitasdeoliveira92@gmail.com

¹¹Bolsista Capes, MS, larissagiroto@yahoo.com.br

¹²Bolsista Fapesp, PhD, suzanatiemi@yahoo.com.br

¹³Pesquisador, PhD, Universidade Estadual Paulista, doug@rc.unesp.br

¹⁴Pesquisador, PhD, Embrapa Café, filipe.pereira@embrapa.br

RESUMO: O objetivo deste trabalho foi explorar os recursos genéticos de genótipos de *Coffea arabica* L. do centro de origem para identificar regiões genômicas associadas à resistência a mancha aureolada. Para isto, 101 genótipos de *C. arabica* da coleção da Etiópia FAO e 11 cultivares foram sequenciados e genotipados através da técnica de genotipagem por sequenciamento (GBS). O GBS resultou em 6.210.920 tags que foram alinhadas ao genoma de *C. arabica* e resultou na identificação de 10.650 SNPs (*Single Nucleotide Polymorphism*). O mesmo material vegetal também foi avaliado fenotipicamente para resistência a mancha aureolada, causada pela *Pseudomonas syringae* pv. *garcae*. A fenotipagem varia entre altamente imune à altamente suscetível. Com esses dados, utilizando o programa FASTmrEMMA foi feito um Estudo de Associação Genômica Ampla. Como resultado, 5 SNPs, com alto potencial para Seleção Assistida, foram associados à resistência a mancha aureolada. Além disso, foi possível identificar genes próximos aos SNPs relacionados ao mecanismo de defesa das plantas. Entre esses estão genes que codificam as proteínas serina/treonina quinase e a CC-NB-LRR.

PALAVRAS-CHAVE: mancha aureolada do cafeeiro, gene R, seleção assistida.

GENOME WIDE ASSOCIATION STUDY FOR *Coffea arabica* L. AND *Pseudomonas syringae* pv. *garcae* INTERACTION

ABSTRACT: The aim of this work was to explore the genetic resources of *Coffea arabica* L. genotypes, from its center of origin, in order to identify genomic regions associated to resistance to bacterial halo blight. For that, 101 Ethiopian *C. arabica* genotypes and 11 cultivars were genotyped using GBS technique (Genotyping by Sequencing). The GBS resulted in 6.210.920 tags that were aligned to the *C. arabica* genome and 10.650 SNPs (Single Nucleotide Polymorphism) were selected. The same plant collection was phenotypically evaluated for resistance to bacterial halo blight that is caused by *Pseudomonas syringae* pv. *garcae*. The phenotyping varies between highly immune to highly susceptible. Using the program FASTmrEMMA a Genomic Wide Association Study was performed. As a result, 5 SNP, with high assisted selection potential, were associated with bacterial halo blight. In addition, it was possible to identify candidate genes, in the proximity of the SNPs associated, related to plant defense, such as serine/threonine kinase and CC-NB-LRR.

KEY WORDS: bacterial halo blight, R gene, assisted selection.

INTRODUÇÃO

A incidência de doenças é um dos principais fatores de redução da produção cafeeira. Entre as doenças bacterianas temos a mancha aureolada, causada pela *Pseudomonas syringae* pv. *garcae* (AMARAL; TEIXEIRA; PINHEIRO, 1956), a qual vem se espalhando pelos estados de São Paulo, Paraná e Minas Gerais (RODRIGUES et al., 2017; ZOCCOLI; TAKATSU; UESUGI, 2011). A doença é predominante em regiões de clima moderado, expostas ao vento e sujeitas a longos períodos de alta umidade relativa (SERA, 2001). Levando - se em conta que o manejo da mancha aureolada é baseado em químicos e quebra-ventos este se torna difícil e com alto custo.

Dessa forma, uma das melhores formas de controle da doença é a obtenção de cultivares resistentes. Neste contexto, o uso de marcadores moleculares pode contribuir para acelerar a seleção de genótipos. Uma forma de identificar

potenciais marcadores moleculares é através do Estudo de Associação Genômica Ampla (GWAS) o qual relaciona polimorfismos, como os SNPs, as características de interesse. Porém, GWAS depende de diversidade genética entre genótipos e de grande número de marcadores para ser realizado com sucesso. Em *C. arabica*, é possível encontrar esse tipo de diversidade em uma coleção de genótipos provenientes da Etiópia (SANT'ANA et al., 2018; SILVESTRINE et al., 2007; FAO, 1968). Assim, a partir dessa coleção e cultivares foi possível fazer um mapeamento de associação para a caracterização de SNPs e de regiões gênicas que atuam na resistência a mancha aureolada.

MATERIAL E MÉTODOS

No presente estudo foram utilizados 101 genótipos de *C. arabica* coletados por uma missão da FAO (FAO, 1968) na Etiópia, as cultivares IPR 99, 100, 101, 102, 103, 104, 105, 107, Catuaí vermelho, IAPAR 59 e Tupi IAC 1669-33. Esses 101 genótipos de *C. arabica* Etiópia fazem parte de uma coleção mantida no campo do IAPAR (23°23'00"S e 51°11'30"W) e possuem uma grande variabilidade genética para resistência a mancha aureolada. Essa variabilidade varia de uma escala de 0 a 5. Onde 0 corresponde a genótipos altamente resistentes a mancha aureolada, 1 à resistente, 2 à moderadamente resistentes, 3 à moderadamente suscetíveis, 4 à suscetíveis e 5 à altamente suscetíveis (MOHAN; CARDOSO; PAIVA, 1978). O material vegetal foi sequenciado através da técnica GBS em parceria com a *Genomic Diversity Facility LIMS*, em Cornell University (Ithaca, NY – EUA), utilizando o Illumina HiSeq2000, *single end*.

O GBS resultou em 6.210.920 tags que foram alinhadas ao genoma de referência de *C. arabica* (ACGC - *Arabica Coffee Genome Consortium*). Através do TASSEL-GBS *pipeline* (GLAUBITZ et al., 2014) foi possível detectar 159.000 SNPs. Com base nesses resultados, utilizando o software TASSEL versão 5.2.37 (BRADBURY et al., 2007), foi feita uma filtragem dos SNPs com os parâmetros de frequência do menor alelo maior que 0.05 (MAF >0.05), Call Rate >0.8 e heterozigosidade < 0.9 que resultou em 10.650 SNPs. A associação dos dados fenotípicos e genotípicos foi feita através do modelo estatístico “*fast multi-locus random-SNP-effect EMMA*” (FASTmrEMMA), desenvolvido dentro do pacote R (WEN et al., 2017). Neste modelo estatístico utilizou-se o p-valor crítico = 0.005 e em um segundo passo o LOD score crítico = 3.

Também foi feita uma busca por genes, no genoma de *C. arabica*, próximos aos SNPs associados à interação com a mancha aureolada.

RESULTADOS E DISCUSSÃO

A partir do método estatísticos FASTmrEMMA foi possível identificar 5 SNPs associados à resistência a mancha aureolada. Desses, 2 encontram-se no cromossomo 2 do subgenoma de *C. eugenoides*. Os outros 3 no Chr 0. O Chr 0 corresponde a um conjunto de regiões sequenciadas que não tiveram posição definida na montagem do genoma.

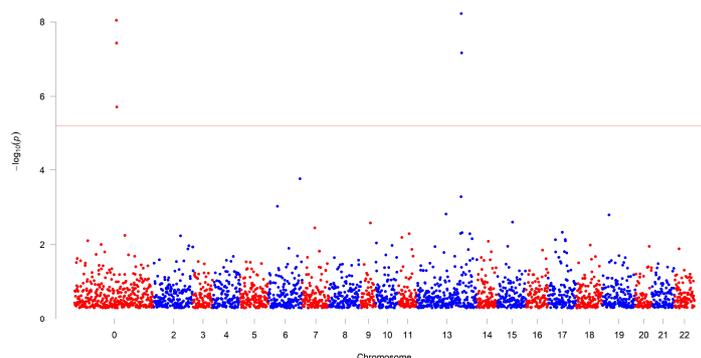


Figura 1. Manhattan plot do GWAS para interação entre *C. arabica* e *Pseudomonas syringae* pv. *garcae*. O plot é baseado em $-\log_{10}(p)$. A linha vermelha indica o limiar para considerar a significância da associação (correção de Bonferroni). No eixo x a numeração de 1 a 11 correspondem aos cromossomos do subgenoma de *C. canephora* e 12 a 22 aos cromossomos do subgenoma de *C. eugenoides*.

Os SNPs associados podem ser detectados em indivíduos, de populações de melhoramento, através de sondas alelo específicas do tipo Taqman®. Dessa forma, em comparação à inoculações artificiais, plantas resistentes podem ser identificadas de forma mais rápida, barata e confiável.

A busca de genes próximos aos SNPs associados revelou genes envolvidos em dois mecanismos de respostas de defesa das plantas. Para os SNPs associados no Chr 0 foi encontrado um gene que codifica a proteína serina/treonina quinase, a qual participa da primeira classe de resposta inata imune da planta. Essa primeira classe é conhecida como *Mamp Triggred Immunity* (MTI), ou seja, é responsável por reconhecer o patógeno através dos *microbe-associated molecular patterns*. Já os SNPs no Chr 2 do subgenoma de *C. eugenoides* estão próximos a um gene que codifica a

proteína CC-NBS-LRR. Essa proteína é responsável por ativar a segunda classe de resposta inata imune, a *Effector Triggered Immunity* (ETI). A ativação da ETI acontece através do reconhecimento de efetores específicos do patógeno. A ETI é um tipo de resposta mais robusta frente ao patógeno e, geralmente, desencadeia morte celular (PENG et al., 2018).

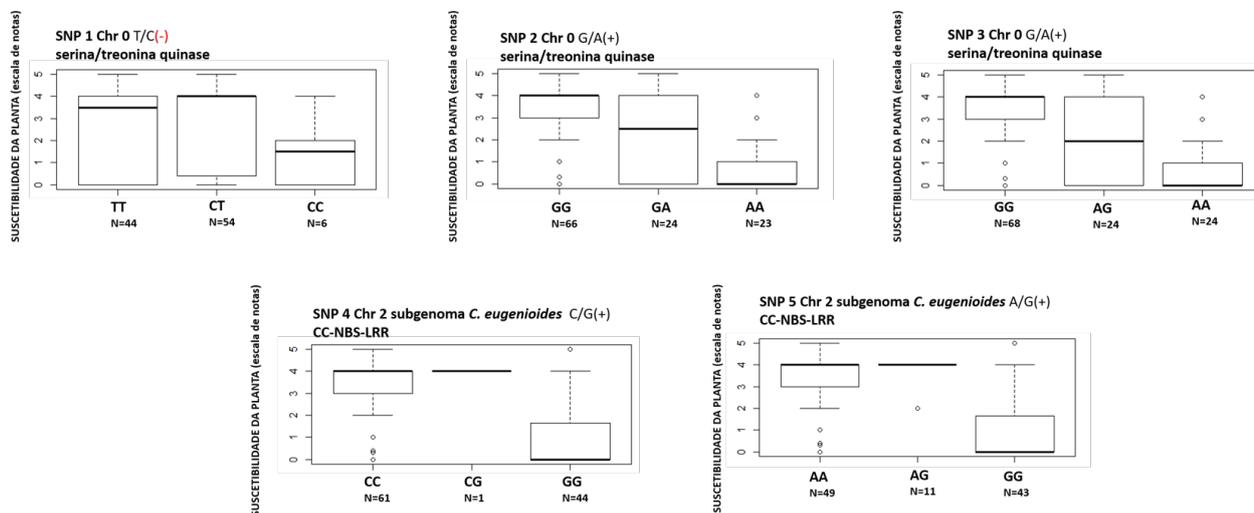


Figura 2. Boxplots dos SNPs associados à interação de *C. arabica* com mancha aureolada. Acima de cada boxplot está a localização do SNP no genoma de *C. arabica*, o nucleotídeo considerado padrão e o nucleotídeo considerado SNP, e entre parênteses se essa associação foi considerada como positiva ou negativa em relação à resistência. Para cada boxplot está descrito o gene próximo ao SNP. No eixo y está a escala de suscetibilidade da planta frente ao patógeno, que varia de altamente imune (0) à altamente suscetível (5). No eixo x está o número de indivíduos que se comporta de forma homocigoto para o alelo padrão, heterocigota e homocigota para o alelo considerado como SNP.

CONCLUSÕES

1. O método estatístico FASTmrEMMA foi eficiente para associar SNPs para a interação de *C. arabica* e mancha aureolada
3. Os SNPs associados possuem potencial para serem usados em Seleção Assistida.
4. Os indivíduos altamente imunes a mancha aureolada devem ser capazes de desencadear tanta MTI quanto ETI.

REFERÊNCIAS BIBLIOGRÁFICAS

- AMARAL, J. F.; TEIXEIRA, C. G.; PINHEIRO, E. D. O bactério causador da mancha-aureolada do cafeeiro. Arquivos do Instituto Biológico, 23, p. 151-155, 1956.
- BRADBURY, P. J. et al. TASSEL: Software for association mapping of complex traits in diverse samples. Bioinformatics, v. 23, n. 19, p. 2633–2635, 2007.
- FAO – Food and Agriculture Organization. Coffee mission to Ethiopia 1964–65. FAO, Rome, 1968.
- GLAUBITZ, J. C. et al. TASSEL-GBS: A high capacity genotyping by sequencing analysis pipeline. PLoS ONE, v. 9, n. 2, 2014.
- MOHAN, S. K.; CARDOSO, R. M. L.; PAIVA, M. A. Resistência em germoplasma de *Coffea* ao cretamento bacteriano incitado por *Pseudomonas garcae*. Pesquisa Agropecuária Brasileira, v. 13, n.1, p. 53-64, 1978.
- PENG, Y.; WERCH, R VAN.; ZHANG, Y.; Convergent and Divergent Signaling in PAMP – Triggered Immunity and Effector – Triggered Immunity. Molecular Plant -Microbe Interactions, v. 31, n. 4, p. 403 – 409, 2018.
- RODRIGUES, L. M. R. et al. First report of mixed infection by *Pseudomonas syringae* pathovars *garcae* and *tabaci* on coffee plantations. Bragantia, v. 76, n. 4, p. 543–549, 2017.
- SANT’ANA, G. C. et al. Genome-wide association study reveals candidate genes influencing lipids and diterpenes contents in *Coffea arabica* L. Scientific Report, v. 8, 2018.
- SERA T. Coffee Genetic Breeding at IAPAR. Crop Breeding and Applied Biotechnology, v. 1, p. 179–199, 2001.
- SILVESTRINI, M. et al. Genetic diversity and structure of Ethiopian, Yemen and Brazilian *Coffea arabica* L. accessions using microsatellites markers. Genetic Resources and Crop Evolution, v. 54, n. 6, p. 1367–1379, 2007.
- ZOCOLI, D. M.; TAKATSU, A.; UESUGI, C. H. Ocorrência de mancha aureolada em cafeeiros na Região do Triângulo Mineiro e Alto Paranaíba. Bragantia, v. 70, n. 4, p. 843–849, 2011.

WEN, Y.-J. et al. Methodological implementation of mixed linear models in multi-locus genome-wide association studies. *Briefings in Bioinformatics*, n. September, p. 1–13, 2017.