

TIAGO VIEIRA SOUSA

**DIVERSIDADE GENÉTICA, GANHOS COM SELEÇÃO E  
SELEÇÃO GENÔMICA AMPLA NA ESPÉCIE *Coffea arabica***

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento, para obtenção do título de Doctor Scientiae.

VIÇOSA  
MINAS GERAIS - BRASIL  
2017

**Ficha catalográfica preparada pela Biblioteca Central da Universidade  
Federal de Viçosa - Câmpus Viçosa**

T

Sousa, Tiago Vieira, 1985-  
S725d           Diversidade genética, ganhos com seleção e seleção  
2017           genômica ampla na espécie *Coffea arabica* / Tiago Vieira Sousa.  
– Viçosa, MG, 2017.  
                  xvii, 116f. : il. (algumas color.) ; 29 cm.

Orientador: Eveline Teixeira Caixeta.  
Tese (doutorado) - Universidade Federal de Viçosa.  
Inclui bibliografia.

1. Café - Melhoramento genético. 2. *Coffea arabica*.  
3. Diversidade genética. 4. Marcadores genéticos.  
I. Universidade Federal de Viçosa. Departamento de Genética e  
Melhoramento. Doutorado em Genética e Melhoramento.  
II. Título.

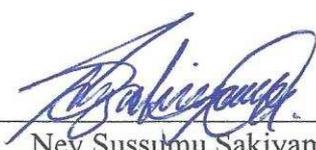
CDD 22 ed. 633.732

TIAGO VIEIRA SOUSA

**DIVERSIDADE GENÉTICA, GANHOS COM SELEÇÃO E  
SELEÇÃO GENÔMICA AMPLA NA ESPÉCIE Coffea arabica**

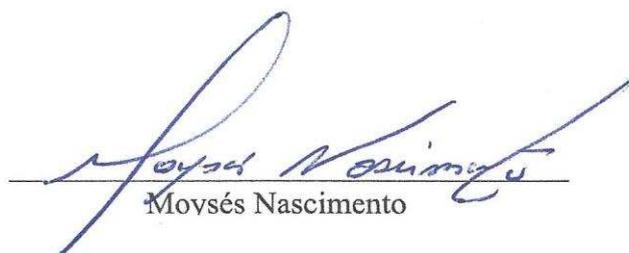
Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento, para obtenção do título de Doctor Scientiae.

APROVADA: 28 de agosto de 2017.



---

Nev Sussumu Sakiyama



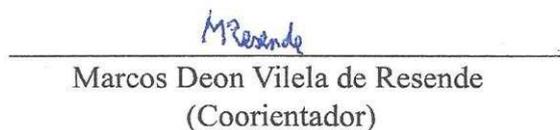
---

Moysés Nascimento



---

Gustavo César Sant'Ana



---

Marcos Deon Vilela de Resende  
(Coorientador)



---

Eveline Teixeira Caixeta  
(Orientadora)

*“Seja a mudança que você deseja ver no mundo”.*

**Mahatma Gandhi**

Dedico a Deus, sem Ele nada seria. Aos meus pais e avós pelo amor e dedicação. E a minha amada esposa, ajudadora e doce companheira.

## AGRADECIMENTOS

A Deus pelo dom da vida. Por tudo o quanto tem me proporcionado.

Agradeço a Universidade Federal De Viçosa (UFV) e ao Curso de Pós-Graduação em Genética e melhoramento pela oportunidade e pelo ensino de excelência.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pela bolsa de estudo.

À Dra. Eveline Teixeira Caixeta, pela orientação, por toda paciência, ajuda e conhecimentos transmitidos.

Aos meus Coorientadores Dr. Antônio Carlos Baião de Oliveira e Dr. Marcos Deon Vilela de Resende, pelos valiosos conhecimentos e sugestões transmitidos.

Ao Dr. Antonio Alves Pereira, pela paciência, amizade e, sobretudo, pelo exemplo de vida e amor à pesquisa com o cafeeiro.

À Dra. Eunize Maciel Zambolim e ao Dr. Laercio Zambolim, pelo auxílio na condução dos trabalhos e pela amizade.

Aos Drs. Ney Sussumu Sakiyama, Moysés Nascimento, Gustavo César Sant'Ana e Luiz Filipe Protasio Pereira pelo aceite ao convite de participação da banca de tese e, principalmente, pelas contribuições valiosas.

Agradeço aos meus pais, Jair e Vilma, de todo o coração, pelo carinho, apoio, confiança, por terem acreditado em mim, sei que não teria chegado até aqui sem vocês.

Aos meus amados irmãos e família por todo amor, carinho e amizade.

À minha esposa amada, Emilly, por ser tão especial em minha vida. Uma companheira, ajudadora, que sempre me motiva e me faz querer algo a mais. Por todo amor e paciência.

Ao amigo, Cristiano de Oliveira, pela ajuda na organização dos dados moleculares.

Aos amigos do Laboratório BioCafé, pela eterna amizade, pelos conselhos e por dividir comigo grandes momentos. Em especial agradeço aos amigos Geleta, Juan, Dani, Júlia, Letícia e Franciele, pelos momentos compartilhados e por tudo que vivemos.

Aos funcionários da UFV e EPAMIG, em especial ao senhor Mário, senhor Paulo, Marco Túlio e Odilon.

## BIOGRAFIA

TIAGO VIEIRA SOUSA, filho de Jair Vieira Sousa e Vilma da Consolação Sousa, nasceu no dia 03 de outubro de 1985, em Itamarandiba, Estado de Minas Gerais.

Concluiu o ensino médio em 2005, na Escola Estadual São João Batista em Itamarandiba-MG.

Em 2007, ingressou no Curso de Agronomia pela Universidade Estadual de Montes Claros (UNIMONTES). Em dezembro de 2011 diplomou-se Engenheiro Agrônomo na referida instituição de ensino, sob orientação da Dra. Márcia Regina Costa.

Em março de 2012, iniciou o curso de mestrado no Programa de Pós-Graduação em Genética e Melhoramento pela Universidade Federal de Viçosa (UFV), sob orientação da Dra. Eveline Teixeira Caixeta, pesquisadora da Embrapa Café. Em julho de 2013 foi sua defesa de dissertação.

Em agosto de 2013, iniciou o curso de doutorado no Programa de Pós-Graduação em Genética e Melhoramento, na Universidade Federal de Viçosa (UFV), sob orientação da Dra. Eveline Teixeira Caixeta, submetendo-se a defesa de tese em agosto de 2017.

## SUMÁRIO

LISTA DE TABELAS .....	ix
LISTA DE FIGURAS .....	xi
RESUMO .....	xiv
ABSTRACT .....	xvi
Introdução geral.....	1
Referências bibliográficas .....	3

### CAPÍTULO 1

Population structure and genetic diversity of coffee progenies derived from Catuaí and Híbrido de Timor revealed by genome-wide SNP marker .....	6
ABSTRACT .....	7
Introduction .....	8
Material and methods .....	10
Genetic material.....	10
Genomic DNA extraction.....	11
Probes design and SNP identification .....	12
Quality analysis of SNP molecular markers.....	13
Genetic-statistical analyses.....	14
Results .....	16
Identification and analysis of SNP markers .....	16
Genetic-statistical analyses.....	16
Allelic frequency, genotypic frequency, and PIC .....	16
Study of population structure .....	17
Genetic distance between pairs of genotypes.....	17
Clustering Analysis .....	18
Principal Coordinate Analysis (PCoA) .....	19
DISCUSSION.....	20
Identification and quality analysis of SNP markers .....	20
Genetic-statistical analyses.....	21
Allele frequency, genotype frequency and PIC.....	21
Study of population structure .....	22

Genetic distance between pairs of genotypes .....	23
Clustering Analysis .....	26
Principal Coordinate Analysis (PCoA) .....	28
REFERENCES .....	43

## **CAPÍTULO 2**

Seleção de progênies de Coffea arabica via procedimento REML/BLUP .....	52
RESUMO .....	53
Introdução .....	54
Material e métodos .....	56
Condução experimental .....	56
Material genético .....	57
Avaliações Fenotípicas .....	57
Análises Estatísticas .....	58
Índice de seleção, correlações genéticas e diversidade genética .....	61
Resultados e discussão .....	62
Análise de deviance .....	62
Componentes de variância e parâmetros genéticos .....	63
Correlação genética .....	66
Diversidade genética.....	68
Índice de seleção.....	69
Conclusão .....	70
Referências Bibliográficas.....	80

## **CAPÍTULO 3**

Seleção genômica ampla em população de Coffea arabica.....	85
RESUMO .....	86
Introdução.....	87
Material e métodos .....	89
Condução experimental .....	89
Material genético .....	90
Avaliações Fenotípicas .....	90
Análises fenotípicas .....	91

Extração do DNA genômico.....	92
Análise de qualidade de marcadores moleculares .....	92
Validação Cruzada.....	93
Método de seleção genômica.....	93
Capacidade preditiva, acurácia da GWS e $n_{QTL}$ .....	94
Densidade de marcadores .....	94
Eficiência da GWS .....	95
Resultados e discussão .....	95
Perspectivas futuras da utilização da GWS em <i>C. arabica</i> .....	101
Conclusão .....	102
Referências Bibliográficas.....	109

## LISTA DE TABELAS

### CAPÍTULO 1

Table 1. Rust resistant and susceptible <i>Coffea arabica</i> parents .....	30
Table 2. <i>Coffea arabica</i> hybrid genotypes and their genealogy .....	31
Table 3. Genotypes selected from the 13 <i>Coffea arabica</i> progenies and their genealogy ..	32
Table 4. <i>Coffea arabica</i> accessions belonging to the germplasm bank of UFV/EPAMIG.	33
Table 5. Number of SNP molecular markers by chromosome after filter 3, genotypic and allelic mean frequencies, and mean PIC of the SNP on each chromosome .....	34
Table 6. The 10 greatest and the 10 smallest genetic distances between the pairs of Arabica coffee genotypes obtained with the SNP analysis .....	35
Table 7. Genetic distances between pairs of parents Catuaí (A, C and E) and Híbrido de Timor (B, D and F) .....	36
Table 8. The greatest and the smallest genetic distances between pairs of individuals within each group.....	37

### CAPÍTULO 2

Tabela 1. Progenies de <i>Coffea arabica</i> avaliadas em Viçosa (MG).....	72
Tabela 2. Características fenotípicas avaliadas nos anos de 2014, 2015 e 2016 em Viçosa (MG) .....	73
Tabela 3. Análise de deviance (ANADEV) e significâncias do teste de razão da verossimilhança (LRT) referentes à análise conjunta de 14 características morfológicas medidas nos anos de 2014, 2015 e 2016, no experimento de progenies de <i>Coffea arabica</i> em Viçosa, MG.....	74

Tabela 4. Estimativa dos parâmetros genéticos referentes à análise conjunta de 14 características morfológicas medidas nos anos de 2014, 2015 e 2016, no experimento de progênies de Coffea arabica em Viçosa, MG. ....	75
Tabela 5. Estimativa das correlações genéticas entre características referentes à análise conjunta de 14 características morfológicas medidas nos anos de 2014, 2015 e 2016, no experimento de progênies de Coffea arabica em Viçosa, MG.....	76
Tabela 6. Índice de seleção, rank, rank-médio e ganho de seleção em porcentagem das progênies de cafeeiros obtidos a partir das características com efeito entre progênies significativo pela análise ANADEV, conforme a Tabela 4.....	79

### **CAPÍTULO 3**

Tabela 1. Progênies de Coffea arabica avaliadas em Viçosa (MG).....	104
Tabela 2. Características fenotípicas avaliadas nos anos de 2014, 2015 e 2016 em Viçosa (MG). ....	105
Tabela 3. Estimativa de parâmetros genéticos obtidos por meio de análises de modelos mistos (REML/BLUP), resultados da análise da Seleção Genômica Ampla (GWS) e estimativa do número de indivíduos a serem avaliados para se obter determinada acurácia seletiva para dezoito características morfoagronômicas em população de melhoramento de Coffea arabica. ....	106
Tabela 4. Acurácia seletiva estimada a partir de diferentes densidades de marcadores SNP e eficiência da seleção genômica ampla (GWS) em relação à seleção fenotípica em população de melhoramento de Coffea arabica. ....	108

## LISTA DE FIGURAS

### CAPÍTULO 1

- Fig. 1 Heredogram of the parents, hybrids and of the 13 progenies that compose the *Coffea arabica* breeding program analyzed in this study; Tables 1, 2 and 3 present the relation of the respective genotypes corresponding to each letter and number;  $\otimes$  = Selfing generation. .... 38
- Fig. 2 SNP molecular markers distributed throughout the UNIGENES from the EST sequences of *Coffea arabica* and of the 11 *Coffea canephora* chromosomes and chromosome "0" of *Coffea canephora*. Chromosome "0" is just a pool of non-ordered sequence scaffolds (Denoëud et al. 2014). .... 39
- Fig. 3 Bar graphs of the InStruct software used to determine the population structure of the 72 *Coffea arabica* genotypes, showing the formation of two groups ( $k = 2$ ); Tables 1 and 2 list the genotypes corresponding to each letter; SUS = Coffee rust susceptible genotype; RES = Coffee rust resistant genotype; BCr = Backcross of  $F_1$  hybrid with resistant recurrent parent; BCs = Backcross of  $F_1$  hybrid with recurrent susceptible parent;  $F_2$  = generation obtained by the selfing of  $F_1$  hybrids..... 40
- Fig. 4 Dendrogram obtained by the UPGMA technique based on the dissimilarity matrix of the arithmetic complement of the weighted index from 72 *Coffea arabica* genotypes; The list of the genotypes corresponding to each letter, from A to K, are presented in Tables 1 and 2; SUS = coffee rust susceptible genotype ; RES = coffee rust resistant genotype; BCr = Backcross of  $F_1$  hybrid with recurrent resistant parent; BCs = Backcross of  $F_1$  hybrid with recurrent susceptible parent;  $F_2$  = generation obtained by the selfing of  $F_1$  hybrids..... 41
- Fig. 5 Principal Coordinates Analysis (PCoA) of the 72 *Coffea arabica* genotypes; a) Groups formed according to type of generation; b) Groups formed according to analysis in InStruct software; c) Groups formed according to dendrogram analysis; P-SUS = Parents of the

Catuaí group; P-RES = Parents of Híbrido de Timor; F<sub>1</sub> = Hybrid obtained by the cross between rust-resistant and rust-susceptible genotypes; F<sub>2</sub> = generation obtained by the selfing of F<sub>1</sub> hybrids; BCr = Backcross of F<sub>1</sub> hybrid with recurrent resistant parent; BCs = Backcross of F<sub>1</sub> hybrid with recurrent susceptible parent; DCr = Resistant Coffee differential clones of *Hemileia vastatrix* races; DCs = Susceptible Coffee differential clones of *Hemileia vastatrix* races; The list of the genotypes corresponding to each letter, from A to K, are presented in Tables 1 and 2. .... 42

## CAPÍTULO 2

Figura 1. Heredograma das 13 progênies de *Coffea arabica* oriundas de cruzamentos entre genitores Catuaí e Híbrido de Timor; C<sub>1</sub>, C<sub>2</sub> e C<sub>3</sub> = genótipos Catuaí amarelo IAC 30, IAC 86 e IAC 64, respectivamente; HdT<sub>1</sub>, HdT<sub>2</sub> e HdT<sub>3</sub> = genótipos Híbrido de Timor UFV 445-46, UFV 440-10 e UFV 530, respectivamente; H<sub>1</sub>, H<sub>2</sub>, H<sub>3</sub>, H<sub>4</sub> e H<sub>5</sub> = híbridos oriundos de cruzamentos entre genitores Catuaí amarelos e Híbrido de Timor; 1, 3, 5 e 7 = progênies de geração de retrocruzamento resistente; 2, 4, 6, 8 e 9 = progênies de geração de retrocruzamento suscetível; 10, 11,12 e 13 = progênies em geração F<sub>2</sub>..... 71

Figura 2. Dendrograma obtido pela técnica UPGMA (Unweighted Pair Group Method using Arithmetic averages) com base na matriz de distância genética euclidiana média a partir de 228 genótipos constituintes das progênies de *Coffea arabica* avaliada, conforme apresentado na tabela 1. CATUAÍ = Catuaí Vermelho IAC 44; OEIRAS = Oeiras MG 6851; 1, 3, 5 e 7 = progênies de geração de retrocruzamento resistente; 2, 4, 6, 8 e 9 = progênies de geração de retrocruzamento suscetível; 10, 11,12 e 13 = progênies em geração F<sub>2</sub>..... 77

Figura 3. Dispersão gráfica das coordenadas principais obtida com os dados da matriz de distância genética euclidiana média a partir de 228 genótipos constituintes das progênies de

Coffea arabica avaliada, conforme apresentado na tabela 1. CATUAÍ = Catuaí Vermelho IAC 44; OEIRAS = Oeiras MG 6851.....	78
--	----

### CAPÍTULO 3

Figura 1. Heredograma das 13 progênies de Coffea arabica oriundas de cruzamentos entre genitores do grupo de cultivares Catuaí e Híbrido de Timor (HdT); C<sub>1</sub>, C<sub>2</sub> e C<sub>3</sub> = genótipos Catuaí amarelo IAC 30, IAC 86 e IAC 64, respectivamente; HdT<sub>1</sub>, HdT<sub>2</sub> e HdT<sub>3</sub> = genótipos Híbrido de Timor UFV 445-46, UFV 440-10 e UFV 530, respectivamente; H<sub>1</sub>, H<sub>2</sub>, H<sub>3</sub>, H<sub>4</sub> e H<sub>5</sub> = híbridos oriundos de cruzamentos entre genitores Catuaí amarelos e Híbrido de Timor; 1, 3, 5 e 7 = progênies de geração de retrocruzamento resistente; 2, 4, 6, 8 e 9 = progênies de geração de retrocruzamento suscetível; 10, 11,12 e 13 = progênies em geração F<sub>2</sub>.....

Figura 2. Marcadores moleculares SNP distribuídos ao longo dos UNIGENES proveniente das sequencias de ESTs de Coffea arabica e dos 11 cromossomos e do “cromossomo 0” de Coffea canephora. “Cromossomo 0” é apenas um conjunto de scaffolds de sequência não ordenada (Denoëud et al. 2014).....

## RESUMO

SOUSA, Tiago Vieira, D.Sc., Universidade Federal de Viçosa, agosto de 2017. **Diversidade genética, ganhos com seleção e seleção genômica ampla na espécie *Coffea arabica*.** Orientadora: Eveline Teixeira Caixeta. Coorientadores: Marcos Deon Vilela de Resende e Antonio Carlos Baião de Oliveira.

A espécie *Coffea arabica* é a mais importante, economicamente, do gênero *Coffea*. Essa espécie é autógama e apresenta base genética estreita. Assim, o sucesso dos programas de melhoramento para essa cultura é desafiador. Nesse sentido, métodos seletivos robustos e precisos são necessários para maximizar os ganhos com seleção, mantendo a variabilidade genética presente nas populações. O uso de marcadores moleculares, por permitir o acesso à informação do DNA dos indivíduos, e da seleção baseada em dados fenotípicos por meio de procedimentos de modelos mistos (REML/BLUP), por permitirem a estimação de parâmetros genéticos de forma acurada mesmo em situação de experimentos desbalanceados, têm se mostrado vantajosos. Finalmente, os métodos de seleção genômica ampla (GWS), que buscam conhecer as associações entre os dados genotípicos e os fenotípicos, tem se mostrado acurado e promissor para diversas espécies vegetais, inclusive para espécies perenes. O objetivo do trabalho foi identificar marcadores SNP e validá-los em estudos de diversidade genética; estimar os parâmetros genéticos e os ganhos com a seleção utilizando a metodologia de modelos mistos (REML/BLUP); aplicar o princípio da GWS e avaliar sua eficiência em população de *C. arabica*. A população em estudo, inicialmente, foi composta por 72 genótipos nos quais os marcadores SNP foram identificados e validados. A seleção fenotípica, por meio do procedimento REML/BLUP, foi realizada em 228 genótipos de cafeeiros. A GWS foi realizada em população composta por 195 genótipos, genotipados com 21.211 e fenotipados para 18 características agronômicas. Um total de 40.000 sondas específicas foram construídas, sendo identificados 91.517 marcadores SNP em 72 indivíduos de *C. arabica*. Após análises de qualidade, 11.187 SNP foram selecionados e validados em análises de diversidade e estrutura genética de

populações. Os marcadores foram eficientes na avaliação da diversidade e estrutura genética de *C. arabica*. Com base nos dados fenotípicos por meio de análises de modelos mistos foi possível estimar os parâmetros genéticos da população avaliada e obter ganhos expressivos com seleção. Os resultados da GWS demonstram o potencial dessa metodologia seletiva para o melhoramento de *C. arabica*, por prever com acurácia os valores genéticos genômicos dos indivíduos e possibilitar a redução no tempo necessário para completar o ciclo de seleção, proporcionando ganhos significativos em eficiência seletiva.

## ABSTRACT

SOUSA, Tiago Vieira, D.Sc., Universidade Federal de Viçosa, August, 2017. **Genetic diversity, selection gains and genome wide selection in the *Coffea arabica* species.** Advisor: Eveline Teixeira Caixeta. Co-advisors: Marcos Deon Vilela de Resende and Antonio Carlos Baião de Oliveira.

The *Coffea arabica* species is the most important, economically, of the *Coffea* genus. This species is autogamous and has a narrow genetic base. Thus, the success of breeding programs for this crop is challenging. In this sense, robust and accurate selective methods are necessary to maximize the gains with selection, keeping the genetic variability present in the populations. The use of molecular markers, by allowing access to the DNA information of the individuals, and selection based on phenotypic data through mixed model procedures (REML/BLUP), because they allow the estimation of genetic parameters accurately even in situations of unbalanced experiments, have proved to be advantageous. Finally, genome wide selection (GWS), which seek to know the associations between genotypic and phenotypic data, have been shown to be accurate and promising for several plant species, including for perennial species. The objective of this study was to identify SNP markers and to validate them in studies of genetic diversity; to estimate genetic parameters and selection gains using the mixed model methodology (REML/BLUP); apply the GWS principle and evaluate its efficiency in *C. arabica* population. The population was initially consisted of 72 genotypes in which the SNP markers were identified and validated. Phenotypic selection, using the REML/BLUP procedure, was performed on 228 coffee genotypes. Finally, GWS was performed in a population consisted of 195 genotypes, genotyped with 21,211 and phenotyped for 18 agronomic traits. A total of 40,000 specific probes were constructed, with 91,517 SNP markers identified in 72 *C. arabica* individuals. After quality analyzes, 11,187 SNPs were selected and validated in analyzes of population structure and genetic diversity of coffee progenies derived from Catuaí and Híbrido de Timor. The markers were efficient

in evaluating the diversity and structure genetic of *C. arabica*. Based on the phenotypic data through analyzes of mixed models, it was possible to estimate the genetic parameters of the evaluated population and obtain expressive gains with selection. Finally, GWS results demonstrate the potential of this selective methodology for the improvement of *C. arabica* by accurately predicting the genomic genetic values of the individuals and enabling the reduction in the time required to complete the selection cycle, providing significant gains in selective efficiency.

## **Introdução geral**

O melhoramento genético possibilita a obtenção de ganhos econômicos, sociais e ambientais e, em vegetais, permitiu o desenvolvimento de diversas cultivares mais produtivas e adaptadas. Para a espécie *Coffea arabica*, as cultivares mais produtivas e, conseqüentemente, mais plantadas são pertencentes ao grupo de cultivares Catuaí (Fernandes et al. 2012), destacando-se também pelo alto vigor vegetativo e porte baixo das plantas (Pereira et al. 2010a). Contudo, essas cultivares são suscetíveis às principais doenças dessa cultura, o que evidencia a necessidade de se lançar novas cultivares que reúnam as características agronômicas de interesse das cultivares Catuaí e que sejam superiores em relação aos estresses bióticos e abióticos. No entanto, por apresentar base genética estreita (Lashermes et al. 1999; Cubry et al. 2008; Lashermes et al. 2011; Setotaw et al. 2013; Sousa et al. 2017), a obtenção de tais cultivares de *C. arabica* é tarefa desafiadora.

Para contornar essas limitações, nos últimos anos os derivados de Híbrido de Timor (HdT), híbrido natural do cruzamento entre *C. arabica* e *C. canephora* (Lashermes et al. 1999), tem sido utilizado nos blocos de cruzamentos para obtenção de progênies por facilitarem a introgressão de alelos de interesse de *C. canephora* em *C. arabica*. Este germoplasma é portador de genes que conferem resistência durável à *Hemileia vastatrix* (Diola et al. 2011; Romero et al. 2014; Alkimim et al. 2017), *Colletotrichum kahawae* (Gichimu et al. 2013; Gichimu et al. 2014; Alkimim et al. 2017) e *Meloidogyne exigua* (Bertrand et al. 2001). Além disso, cultivares derivadas de HdT demonstraram superioridade na qualidade de bebida em relação às cultivares tradicionais de café arábica (Pereira et al. 2010b; Sobreira et al. 2015) e tem possibilitado o aumento da base genética de *C. arabica* (Setotaw et al. 2013).

Em busca de cultivares que reúnam o máximo de vantagens agronômicas, os programas de melhoramento de *C. arabica* realizam hibridações entre genitores Catuaí e

HdT, obtendo progênies que são avaliadas e, as melhores, selecionadas. Esta seleção deve ser acurada pois, pode definir se o programa de melhoramento terá sucesso, e em espécies perenes, por apresentar longo período juvenil, isso é intensificado. Nessas espécies, os caracteres fenotípicos podem ser expressar espaçados por longos períodos de dias, meses ou anos. Nesse sentido, é evidente a importância de utilizar métodos seletivos acurados nos programas de melhoramento para o desenvolvimento das cultivares.

Basicamente, os métodos seletivos utilizados nos programas de melhoramento genético têm sido realizados pela avaliação fenotípica dos indivíduos em um local e ano. Assim, apresentam limitações, pois as características fenotípicas podem ser altamente influenciadas pelo ambiente (Hoffmann et al. 2005; Andrew et al. 2010), sobretudo, quando estas são governadas por muitos genes e/ou apresenta baixa herdabilidade. Além disso, em plantas perenes como o café arábica que apresentam longo período juvenil e bienalidade produtiva, a interação genótipo ambiente (colheitas) torna-se fator limitante para esse tipo de seleção fenotípica (Andrew et al. 2010). Dessa forma, métodos seletivos que se baseiam na seleção de progênies avaliadas em mais de um ano e/ou local têm sido mais precisos e têm proporcionado maiores valores de acurácia (Pereira et al. 2013).

Nesse trabalho, objetivou-se identificar marcadores SNP e validá-los em estudos de diversidade genética; estimar os parâmetros genéticos e os ganhos com a seleção utilizando a metodologia de modelos mistos (REML/BLUP); aplicar o princípio da GWS e avaliar sua eficiência em população de *C. arabica*.

## Referências bibliográficas

- Alkimim ER, Caixeta ET, Sousa TV, et al (2017) Marker-assisted selection provides arabica coffee with genes from other *Coffea* species targeting on multiple resistance to rust and coffee berry disease. *Molecular Breeding* 37:6. doi: 10.1007/s11032-016-0609-1
- Andrew RL, Wallis IR, Harwood CE, Foley WJ (2010) Genetic and environmental contributions to variation and population divergence in a broad-spectrum foliar defence of *Eucalyptus tricarpa*. *Annals of Botany* 105:707–717. doi: 10.1093/aob/mcq034
- Bertrand B, Anthony F, Lashermes P (2001) Breeding for resistance to *Meloidogyne exigua* in *Coffea arabica* by introgression of resistance genes of *Coffea canephora*. *Plant Pathology* 50:637–643. doi: 10.1046/j.1365-3059.2001.00597.x
- Cubry P, Musoli P, Legnate H, et al (2008) Diversity in coffee assessed with SSR markers: Structure of the genus *Coffea* and perspectives for breeding. *Genome* 51:50–63. doi: 10.1139/G07-096
- Diola V, de Brito GG, Caixeta ET, et al (2011) High-density genetic mapping for coffee leaf rust resistance. *Tree Genetics & Genomes* 7:1199–1208. doi: 10.1007/s11295-011-0406-2
- Fernandes ALT, Partelli FL, Bonomo R, Golynski A (2012) A moderna cafeicultura dos cerrados brasileiros. *Pesquisa Agropecuária Tropical* 42:231–240. doi: 10.1590/S1983-40632012000200015
- Gichimu B, Gicheru E, Mamati G, Nyende A (2013) Variation and association of cup quality attributes and resistance to Coffee Berry Disease in *Coffea arabica* L. composite cultivar, Ruiru 11. *African Journal of Horticultural Science* 7:22–35.
- Gichimu BM, Gichuru EK, Mamati GE, Nyende AB (2014) Occurrence of Ck-1 gene conferring resistance to Coffee Berry Disease in *Coffea arabica* cv. Ruiru 11 and its parental genotypes. *Journal of Agricultural and Crop Research* 2:51–61.

- Hoffmann AA, Shirriffs J, Scott M (2005) Relative importance of plastic vs genetic factors in adaptive differentiation: geographical variation for stress resistance in *Drosophila melanogaster* from eastern Australia. *Functional Ecology* 19:222–227. doi: 10.1111/j.1365-2435.2005.00959.x
- Lashermes P, Combes M-C, Robert J, et al (1999) Molecular characterisation and origin of the *Coffea arabica* L. genome. *Molecular and General Genetics MGG* 261:259–266. doi: 10.1007/s004380050965
- Lashermes P, Combes MC, Ansaldi C, et al (2011) Analysis of alien introgression in coffee tree (*Coffea arabica* L.). *Molecular Breeding* 27:223–232. doi: 10.1007/s11032-010-9424-2
- Pereira AA, Carvalho GR, Moura WM, et al (2010a) Cultivares: Origem e suas Características. In: Reis PR, Cunha RL (eds) *Café arábica do plantio à colheita*, EPAMIG. Lavras, pp 167–221
- Pereira MC, Chalfoun SM, Carvalho GR de, Savian TV (2010b) Multivariate analysis of sensory characteristics of coffee grains (*Coffea arabica* L.) in the region of upper Paranaíba. *Acta Scientiarum Agronomy*. doi: 10.4025/actasciagron.v32i4.4283
- Pereira TB, Carvalho JPF, Botelho CE, et al (2013) Eficiência da seleção de progênies de café F4 pela metodologia de modelos mistos (REML/BLUP). *Bragantia* 72:230–236. doi: 10.1590/brag.2013.031
- Romero G, Vásquez LM, Lashermes P, Herrera JC (2014) Identification of a major QTL for adult plant resistance to coffee leaf rust (*Hemileia vastatrix*) in the natural Timor hybrid (*Coffea arabica* x *C. canephora*). *Plant Breeding* 133:121–129. doi: 10.1111/pbr.12127
- Setotaw TA, Caixeta ET, Pereira AA, et al (2013) Coefficient of Parentage in *Coffea arabica* L. Cultivars Grown in Brazil. *Crop Science* 53:1237–1247. doi: 10.2135/cropsci2012.09.0541

Sobreira FM, Oliveira ACB de, Pereira AA, et al (2015) Sensory quality of arabica coffee (*Coffea arabica*) genealogic groups using the sensogram and content analysis. *Australian Journal of Crop Science* 9:486–493.

Sousa TVTV, Caixeta ETET, Alkimim ERER, et al (2017) Molecular markers useful to discriminate *Coffea arabica* cultivars with high genetic similarity. *Euphytica* 213:75.  
doi: 10.1007/s10681-017-1865-9

## **CAPÍTULO 1**

**Population structure and genetic diversity of coffee progenies derived from Catuaí  
and Híbrido de Timor revealed by genome-wide SNP marker**

VIÇOSA  
MINAS GERAIS - BRASIL  
2017

## ABSTRACT

The use of Single Nucleotide Polymorphism (SNP) molecular markers has provided advances in selection methodologies used in breeding programs of different crops, reducing cost and time of cultivar release. Despite the great economic and social importance of *Coffea arabica*, studies with SNP markers are scarce and a small number of SNP are available for this species, when compared with other crops of agronomic importance. Thus, the objective of this study was to identify and validate SNP molecular markers for the species *Coffea arabica* and to introduce these markers to genetic breeding by means of an accurate analysis of the diversity and genetic structure of breeding populations of this species. After quality analyses, 11,187 SNP markers were selected from the coffee population obtained from crosses between the genotypes Catuaí and Híbrido de Timor. A great number of markers was distributed in the 11 chromosomes and transcribed, and were used to estimate the genetic dissimilarity among the individuals of the breeding population. Dendrogram analysis and the Bayesian approach demonstrated the formation of two groups and the accurate discrimination of all genotypes evaluated. The expressive number of SNP molecular markers distributed throughout *C. arabica* genome was efficient to discriminate all the accessions evaluated in the experiment, clustering them according to their genealogies. This work identified mixtures within the progenies. It also analyzed in detail the current parents used in studies and identified new parents to be introduced to the breeding program. The study discussed population structure and its consequence in obtaining improved varieties of *C. arabica*.

**Keywords:** *Coffea arabica*; Introgression; Next generation sequence; Genetic relationships; Molecular breeding; InStruct.

## Introduction

Rapid population growth associated with several climate changes and with restricted expansion of agricultural frontiers are challenges for modern agriculture in the continuous food supply. To overcome these limitations, genetic breeding programs must have precise and accurate selective methods to maximize selection gains and maintain the genetic variability of the population (Hallauer 2011).

However, the progress of breeding programs has culminated in the reduction of genetic variability of improved populations (Rodgers et al. 1983; Ortiz et al. 2003; Smith et al. 2004). This problem may be worse in species with narrow genetic base, such as *Coffea arabica*, which belongs to the genus *Coffea*. This genus comprises 104 species (Davis et al. 2006; Davis 2010; Davis 2011), and *C. arabica* and *C. canephora* are the most important agronomic species, corresponding to 63% and 37% of the world's coffee production in 2016, respectively (ICO 2017).

*C. arabica* is allotetraploid, originated from the natural cross between the species *C. eugenioides* and *C. canephora* (Lashermes et al. 1999), and is autogamous with about 10% cross fertilization (Carvalho and Krug 1949). Allied to autogamy, the low number of plants initially distributed throughout the world explains the genetic base of this species (Setotaw et al. 2013). Thus, the selection of superior genotypes by traditional methods is difficult, for they are based exclusively on phenotypic traits. This occurs because genetically distinct individuals may be phenotypically similar, which reduces selective efficiency. To overcome this problem, molecular markers have been used as an auxiliary tool to accurately discriminate genotypes (Ferrão et al. 2015; Sousa et al. 2017) in order to accelerate and make breeding programs feasible (Ceccarelli 2015).

The use of markers is particularly desirable for coffee, a perennial species with a narrow genetic base, due to the time and resources required to release a new cultivar (Ferrão

et al. 2015). The molecular information in studies on genetic structure and diversity of populations allows the knowledge of the genetic structure and similarity between the individuals, and consequently provides precise and accurate selection of superior genotypes, maximizing selection gains (Sousa et al. 2017). By means of molecular markers, several studies have demonstrated the lower genetic diversity of arabica coffee in relation to canephora coffee (Lashermes et al. 1999; Cubry et al. 2008; Lashermes et al. 2011; Ferrão et al. 2015). Thus, the breeding program of EPAMIG, in partnership with UFV and EMBRAPA Café, has aimed at the introgression of genes of interest, present in *C. canephora*, into *C. arabica*. This introgression is facilitated by the use of Híbrido de Timor (HdT), a natural hybrid between the species *C. arabica* and *C. canephora* (Lashermes et al. 1999), in crosses with *C. arabica*. HdT's germplasm carries genes that confer durable resistance to *Hemileia vastatrix* (Diola et al. 2011; Romero et al. 2014; Alkimim et al. 2017), *Colletotrichum kahawae* (Gichimu et al. 2013; Gichimu et al. 2014; Alkimim et al. 2017), and *Meloidogyne exigua* (Bertrand et al. 2001; Pereira et al. 2016). All cultivars released by this breeding program are derived from HdT and have demonstrated superior cup quality when compared with traditional arabica coffee cultivars (Pereira et al. 2010b; Sobreira et al. 2015).

Different molecular markers have been identified and made available for coffee (Rovelli et al. 2000; Combes et al. 2000; Moncada and McCouch 2004; Cubry et al. 2008; Missio et al. 2009a; Vieira et al. 2010; Ferrão et al. 2015). These species-specific markers together with the random markers have been used for different genetic studies in *C. arabica*, such as analyses of genetic diversity and molecular characterization of germplasm (Maluf et al. 2005; Oliveira et al. 2007; Missio et al. 2009b; Missio et al. 2011; Aerts et al. 2013), linkage maps (Pearl et al. 2004; Diola et al. 2011; Pestana et al. 2015; Moncada et al. 2016), marker assisted selection (Alkimim et al. 2017), and cultivars fingerprinting (Sousa et al.

2017). Despite being advantageous, molecular discrimination has some limitations, such as the difficult identification of polymorphic molecular markers distributed throughout the genome of the species. Low percentage of polymorphic markers (10%) is observed for the molecular markers currently available for *C. arabica*, making the selection process difficult and costly (Sousa et al. 2017).

With the advent of modern and robust sequencing platforms, the identification and use of a new type of molecular marker, Single Nucleotide Polymorphism (SNP), has been economically made available. Recently, SNP markers have been identified for several plant and animal species. SNP can be made available on chips, which may present hundreds of thousands of markers distributed throughout the genome of the species. These markers can be associated with the genes that control the main traits of agronomic interest (Heffner et al. 2009). In addition, SNP markers have low cost per data point (Elsik et al. 2009; Matukumalli et al. 2009). These facts have increased the accuracy of the selection methods used in breeding programs for several plant species (Crossa et al. 2010).

To introduce new molecular markers and increase the efficiency of their use in coffee genetic breeding, the present work identified and validated SNP markers distributed throughout the whole *C. arabica* genome. With the large-scale genotyping of the SNP markers obtained in this work, coffee populations were analyzed aiming at the study of genetic diversity and genetic structure of breeding populations.

## **Material and methods**

### **Genetic material**

A population of coffee plants of the *C. arabica* breeding program of EPAMIG/UFV/EMBRAPA was selected for the identification of the SNP markers. The genotypes have been maintained in the experimental area of the Department of Plant Pathology of the Universidade Federal de Viçosa.

The population consisted of 72 coffee genotypes from different generations of crosses between contrasting parents in relation to resistance to coffee rust, and was schematized in the heredogram (Figure 1). The colors dark green, dark blue, dark yellow, light green, light blue and light yellow correspond, respectively, to rust-susceptible parents, rust-resistant parents, F<sub>1</sub> generation, genotypes obtained from the backcross between F<sub>1</sub> hybrid with the recurrent susceptible parent (BCs), genotypes obtained from the backcross between F<sub>1</sub> hybrid with recurrent rust-resistant parent (BCr), and the F<sub>2</sub> generation.

Crosses were carried out between three rust-resistant Híbridos de Timor (HdT) accessions with three rust-susceptible Catuaí varieties (Table 1). Five coffee rust-resistant F<sub>1</sub> hybrids (Table 2) were selected, which were selfed and backcrossed, resulting in 13 arabica coffee progenies (Table 3). In each progeny, four to five individuals were selected. In addition to these genotypes, four accessions of arabica coffee belonging to the germplasm bank of these institutions were selected to compose the study population, in order to increase the genetic variability of the population under analysis (Table 4). Thus, the population was composed of 72 genotypes of coffee plants of different generations.

### **Genomic DNA extraction**

Young and fully expanded leaves from 72 genotypes were collected, and the genomic DNA was extracted using the methodology described by Diniz et al. (2005). DNA concentration was measured using the NanoDrop 2000, and DNA quality was evaluated in a 1% agarose gel.

DNA concentration of the samples was standardized and sent to Rapid Genomics, Florida/USA, for the construction of probes, sequencing and identification of SNP molecular markers.

### **Probes design and SNP identification**

The probes used in this work were developed for the species *C. arabica* and *C. canephora* (Resende et al. 2016). These authors constructed specific probes from reference sequences derived from information contained in the ESTs databases (Vieira et al. 2006) and from the reference genome of *C. canephora* (Denoeud et al. 2014). Probes were constructed in a way that the whole genome was covered, in order to obtain specific probes in transcribed and non-transcribed regions.

The ESTs database is originated from the Brazilian Coffee Genome Project, which contains more than 200 thousand ESTs (Expressed Sequence Tag), and corresponds to about 33 thousand transcripts, known as UNIGENE (Vieira et al. 2006). The *C. canephora* reference genome has an estimated size of 710 Mb, and 25,574 genes have been recorded in this species (Denoeud et al. 2014).

Polymorphic probes (Resende et al. 2016) were obtained as follows: from the sequences deposited in the ESTs database. 10,183 probes of *C. canephora*, 42,732 probes of *C. arabica* and 56,026 probes common to both species were obtained, totaling 108,941 probes. Of this total, 38,266 aligned with *C. canephora* genome. However, 21,176 probes were removed for being monomorphic (zero SNP), resulting in 17,090 probes. Of these, one probe per gene (from the *C. canephora* genome) was selected, which resulted in the selection of 7,347 probes, corresponding to 7,437 genes, where 889 probes presented four SNPs per probe.

The 889 probes for being highly polymorphic, in addition to being aligned with the *C. canephora* genome, probes sequences were extracted from UNIGENE contigs.

1,879 probes were obtained exclusively from ESTs of *C. arabica* that did not align in the reference genome of *C. canephora*. This suggests that these probes are obtained from genes exclusive to *C. arabica*.

From the *C. canephora* genome, 706,227 probes were designed. Of these, 11,879 probes were selected in 11,879 genes. Thus, summed to the 7,437 probes, 19,316 specific probes were selected, corresponding to 19,316 genes. In *C. canephora*, 25,574 genes were estimated (Denoeud et al. 2014).

In addition to probes for gene regions, probes for non-gene regions were selected. 18,006 probes were selected, and the interval between two probes was of 10 kb at most. Probes were distributed in the non-gene regions of the *C. canephora* genome, totaling 40,000 probes.

After capturing 40,000 target regions in the coffee genome, *C. canephora* and *C. arabica* genotypes were sequenced using the Illumina platform. SNPs were identified using bioinformatics tools. This sequencing resulted in the identification of 162,026 SNPs (SNPs specific to *C. arabica*, specific to *C. canephora*, and common to both species) in 27,651 polymorphic probes, with a mean of five SNPs per probe.

The SNP markers identified for *C. arabica* (Resende et al. 2016) were used in the present work.

### **Quality analysis of SNP molecular markers**

Low quality SNP were removed from the data set using minimum criteria for each marker in three quality analyses (filter one, filter two, and filter three). Quality analyses were performed in the VCFtools software (Danecek et al. 2011) and Rbio (Bhering 2017). The quality parameters used in the first filter were MinDP3, DPrange15-750, Miss0.4, MAF0.01, and MinQ10. For the second filter, the quality parameters were MinDP5, DPrange20-350, Miss0.2, MAF0.03 and MinQ20. And finally, the quality parameters used in the third filter were MAF0.1 and CR0.95. In addition, SNP sets that presented identical genotyping results in all individuals evaluated were identified in filter 3. Each set of markers was represented

by an SNP, and the other SNPs were removed from the data set to avoid redundancy and multicollinearity problems.

MinDP is defined as the minimum coverage per SNP and per individual. Values below the critical level determined by this parameter were defined as missing data; DPrange is the mean coverage range of the population for which an SNP is maintained. SNP markers with mean coverage in the population outside the range determined in each filter were removed. Miss refers to data missed after filtering by the MinDP parameter. SNP markers with Miss values above that set in the parameter were removed. MAF is the Minor allele frequency that refers to the frequency at which the alternative allele occurs in a given population. MinQ is the minimum quality accepted for SNP. This parameter is based on the quantity and quality of alignment of the reads, and SNP with quality values lower than those established by this criterion were removed. CR is the minimum proportion of genotypes per marker with non-missing data.

### **Genetic-statistical analyses**

For the genetic-statistical analyses of population data, since the marker was bi-allelic and codominant, the genotypes  $A^1A^1$ ,  $A^1A^2$  or  $A^2A^2$  were coded as 11, 12 and 22, respectively.

Allele and genotype frequencies of the selected SNP markers were estimated. In addition, polymorphism information content (PIC) (Botstein et al. 1980) of each SNP obtained from the following equation was calculated as follows:

$$PIC = 1 - \sum_{i=1}^a p_i^2 - \sum_{i,j=1}^a \sum_{i \neq j}^a p_i^2 p_j^2$$

Where  $p_i$  is the frequency of the  $i$ -th allele of the studied locus.

Population structure was also studied. This analysis was performed by the Bayesian clustering algorithm of the InStruct software (Gao et al. 2007). K values ranged from 1 to 6, with 5 replications. Each run was implemented with a period of 50,000 burn-in, followed by 200,000 MCMC (Markov Chain Monte Carlo).

The genetic dissimilarity matrix was obtained by the arithmetic complement of the weighted index implemented in the GENES software (Cruz 2013). Genetic distance was obtained from the equation:

$$D_{ii'} = 1 - \left( \frac{1}{2} \sum_{j=1}^L p_j c_j \right)$$

where:

$D_{ii'}$  = genetic distance between the pairs of accessions i and i'

$p_j = \frac{a_j}{A}$ : weigh associated with locus j, determined by:

$a_j$ : total number of alleles of locus j;

A: total number of alleles studied;

$$\sum_{j=1}^L p_j c_j = 1$$

$c_j$ : number of common alleles between the pairs of accessions i and i'

The dendrogram was constructed using the UPGMA methodology (Unweighted Pair Group Method using Arithmetic averages) implemented in the MEGA 7.0 software (Kumar et al. 2016).

The principal coordinate analysis (PCoA) was performed by the Genes software (Cruz 2013) using the vegan package implemented in the R software. PCoA is the multivariate statistical analysis where eigenvalues were extracted from the genetic dissimilarity matrix. This analysis is advantageous since it can be applied when the relations between the variables are not linear (Inácio et al. 2002).

## Results

### Identification and analysis of SNP markers

A total of 91,517 SNP markers were identified with the probes distributed throughout the coffee genome. After analyzing the quality of the SNP markers using the first, the second, and the third filter, 34,892, 7,058, and 38,380 SNP were eliminated, respectively. Filter one, filter two, and filter three reduced the original set of SNP markers in 38.13%, 45.84% and 87.78%, respectively. Therefore, after quality analyses, 11,187 SNP markers were selected (Figure 2).

By the analysis of the four datasets (data without a filter, data after filter one, data after filter two, and data after filter three) in the *C. canephora* reference genome and in the database of ESTs of *C. arabica* (UNIGENE), markers were confirmed to be distributed throughout the genome. The greatest number of SNP markers was identified on chromosomes 0 and 1 of *C. canephora*. The "0" chromosome is not a true chromosome; in fact, it is just a set of unsorted sequence scaffolds. The lowest number of SNP markers was observed in the UNIGENE (Figure 2).

### Genetic-statistical analyses

Allelic frequency, genotypic frequency, and PIC

All remaining SNP markers were listed per chromosome and UNIGENE probes, and had the means of allele and genotype frequencies per chromosome estimated (Table 5). The mean allelic frequency was 0.5725 and 0.4275 for alleles  $A^1$  and  $A^2$ , respectively. The mean genotypic frequency was 0.1852, 0.7745, and 0.0402 for genotypes  $A^1A^1$ ,  $A^1A^2$  and  $A^2A^2$ , respectively. The mean PIC per chromosome was also estimated (Table 5). The SNP of chromosomes 1 and 9 had the highest mean PIC values, 0.3526 and 0.3546, respectively. The mean PIC value was 0.3503.

### Study of population structure

To analyze the population structure of the genotypes of interest for breeding in studies, 11,187 SNP markers were used. Two groups ( $K = 2$ ) would be the optimal number (Figure 3).

In the first group, individuals were represented by bars, and the blue color was predominant. In the second group, individuals were represented by bars, and the red color was predominant. Parents of the Catuaí group, the progenies obtained by susceptible backcrosses, and most of the plants of the  $F_2$  progenies were clustered in the first group. The second group clustered the HdT parents (except the parent UFV 530 [F]), all individuals obtained from resistant backcrosses (except one individual of progeny 3 (3-2)), and plants of progenies 10, 11 and 12 ( $F_2$ ). The genotype UFV 530 (F) was genetically intermediate between the commercial cultivars of the Catuaí group and HdT UFV 445,46 (B) and UFV 440-10 (D), accounting for approximately 54% of the blue color bar. Results of this analysis showed the greater similarity of the genotypes obtained by susceptible and resistant backcrosses with the parents of the Catuaí group and HdT, respectively. In addition, all  $F_1$  hybrids (except for H513-5 (K)) had similarity of about 50% with the two groups. Hybrid H513-5 (K) presented intermediate similarity between its parents, UFV 2148-57 (E) and UFV 530 (F). All individuals of progeny 9, obtained from the backcrossing between hybrid H513-5 (K) and UFV 2148-57 (E) also showed similarity between their parents, demonstrating the high discriminatory ability of the SNP markers selected in this study.

### Genetic distance between pairs of genotypes

A genetic distance matrix between the pairs of coffee genotypes was constructed using the 11,187 SNP markers. The greatest genetic distance (0.2542) was observed between

genotypes CIFC 134/4 and UFV 440-10 (D). The smallest genetic distance (0.0673) was observed between genotypes H419-10 (H) and H514-7 (I) (Table 6).

Genetic distances between the pairs of parents used to obtain the study population were evaluated (Table 7). The greatest genetic distance (0.2336) was observed between the parents UFV 2148-57 (E) and UFV 440-10 (D). The smallest genetic distance (0.0836) was observed between the parents UFV 2143-236 (A) and UFV 2154-344 (C). The mean genetic distances between pairs of parents Catuaí and the Híbrido de Timor were 0.0910 and 0.1442, respectively.

The coffee plants population was divided into four groups, based on their generation: 1) F<sub>1</sub> hybrids; 2) F<sub>2</sub> generation; 3) BC<sub>r</sub>; and 4) BCs. The greatest genetic distances were observed between genotypes belonging to the F<sub>2</sub> progenies. The smallest genetic distance was observed between the F<sub>1</sub> hybrids (Table 8).

### Clustering Analysis

A dendrogram was constructed using the UPGMA clustering technique from the values of the distance matrix generated by the arithmetic complement of the weighted index (Figure 4). All genotypes were accurately discriminated. Results of the clustering analysis using the UPGMA methodology were in accordance with those obtained by the Bayesian analysis, forming two large clusters.

The first group clustered the rust-susceptible parents and the rust-resistant parent UFV 530 (K), all individuals from the backcross with the recurrent susceptible parent, all hybrid F<sub>1</sub>, and most individuals of progeny in F<sub>2</sub> generation.

Two large subgroups (1.a and 1.b) were formed in this group. Subgroup 1.a clustered all the rust-susceptible parents, the parent HdT UFV 530 (F), the hybrid F<sub>1</sub> H513-5 (K), plants of progenies 2, 6, 8 and 9 in BCs, progeny 13 in F<sub>2</sub> generation, and three of the four

clones of differential coffee plant hosts for *Hemileia vastatrix* Berk. et Br CIFIC 128/2, CIFIC 635/3, and CIFIC 33/1. Subgroup 1.b clustered four of the five F<sub>1</sub> hybrids, individuals of progenies 10, 11 and 12 in F<sub>2</sub> generation, individuals of progenies 2, 4 and 8 in BCs generation, and two plants of progeny 3 in BCr generation.

The second group clustered two resistant parents, UFV 445-46 (B) and UFV 440-10 (D), four plants in F<sub>2</sub> generation (progenies 10, 11, and 12), and all progenies resistant backcross, except for two plants of progeny 3 (3-2 and 3-3).

One plant of progeny 1 (1-5) and the accessions of the germplasm that correspond to clones of differential coffee plant hosts for *Hemileia vastatrix* Berk. et Br CIFIC 134/4 were not clustered in any of the major groups.

#### Principal Coordinate Analysis (PCoA)

Figure 5 shows the graphical dispersion of the scores obtained from the Principal Coordinate Analysis (PCoA) from the genetic dissimilarity matrix of coffee plants. As in the Bayesian analysis, PCoA showed the genetic distance between the parents of the Catuaí group and the HdT parents. In addition, this analysis demonstrated the genetic dissimilarity between the parent HdT UFV 530 (F) and the other HdT evaluated, and also the genetic dissimilarity between this parent and the F<sub>1</sub> hybrids evaluated in this study. Progenies 9 and 13 and hybrid H513-5 (K) were clustered in the graphic region comprised between their parents, UFV 2148-57 (E) and UFV 530 (F). The same occurred with the other F<sub>1</sub> hybrids and progenies obtained from the other crosses. As demonstrated in the InStruct software, the graphic dispersion showed the highest similarity of the progenies obtained by susceptible and resistant backcrosses with their respective parents, Catuaí and HdT. Moreover, the greatest graphic dispersion was observed among F<sub>2</sub> individuals.

Results of PCoA were in agreement with those of population structure. The region where the four of the five F<sub>1</sub> hybrids were clustered divides the population into two groups, and similar result was observed in the other analyses.

## **DISCUSSION**

### **Identification and quality analysis of SNP markers**

The data set composed of the SNP markers was subjected to quality analysis, which eliminated 80,330 SNP. To ensure that only informative SNP markers were applied to *C. arabica*, the 11,187 SNP markers obtained with more stringent parameters (filter 3) were used in the other analyses. With this strategy, more accurate and safer markers were developed with the potential to generate a lower missing percentage in the populations to be analyzed and greater probability of success in determining the nitrogen bases of the SNP markers (Laurie et al. 2010; Yang et al. 2011).

The SNP markers obtained in this species are distributed on the 11 chromosomes and on chromosome "0", which is a pool of non-ordered sequence scaffolds of *C. canephora*. The identification of SNP in these chromosomes is the result of the construction of probes from the reference genome of *C. canephora* (Denoeud et al. 2014), since the genome of *C. arabica* is not available. The expressive number and distribution of SNP markers throughout the genome of *C. canephora* evidenced the efficiency and precision of the technique used in the construction of polymorphic probes (Resende et al. 2016). The greatest number of SNP markers was identified on chromosomes 0 and 1. This fact may be related to their length, since, for this species, a variation of 2.06 to 5.30  $\mu$ M was observed (Clarindo and Carvalho 2008).

Because *C. arabica* originates from the cross between *C. canephora* and *C. eugenoides* (Lashermes et al. 1999), some *arabica* chromosomes that are exclusively originated from *C. eugenoides* might not be sampled with the probes constructed from the

reference genome of *C. canephora*. To solve this problem, polymorphic probes were constructed from the ESTs sequences (Expressed Sequence Tag), originated from the Brazilian Coffee Genome Project. This database has more than 200,000 ESTs, of which 130,792 are originated from *C. arabica*, which identified approximately 33,000 transcripts (UNIGENE) of the Coffee Genome Project (Vieira et al. 2006). The SNP markers identified from the UNIGENE probes are crucial for *C. arabica* breeding programs for the EST sequences are functional gene sequences, which increases the probability of finding markers associated with genes of interest.

With this large number of informative SNP markers distributed throughout the genome of the species, it is evident the possibility of application of accurate studies on diversity and population structure, as well as selection and broad genomic association in *C. arabica*. In vegetables, this type of analysis has provided selection gains in polygenic traits, with low heritability, which has stimulated the adoption of such procedures (Crossa et al. 2010; Garcia et al. 2011; Heslot et al. 2012; Resende et al. 2012a; Resende et al. 2012b; Azevedo Peixoto et al. 2017).

### **Genetic-statistical analyses**

Allele frequency, genotype frequency and PIC

Polymorphism Information Content (PIC), genotype and allele frequencies for the distinct loci were estimated. The mean PIC value of the 11,187 SNP markers in this population was of 0.3503. The low PIC value evidences the narrow genetic base of *C. arabica*. Similar results were observed in other studies in which PIC values were obtained from SSR and RFLP molecular markers, being multi-allelic and very informative (Lashermes et al. 1999; Poncet et al. 2006; Hendre et al. 2008; Ferrão et al. 2015). Thus, it can be inferred that despite being bi-allelic, the SNP were very informative.

The mean frequency of heterozygote genotypes ( $A^1A^2$ ) was higher than the mean frequencies of homozygote genotypes ( $A^1A^1$  and  $A^2A^2$ ). Despite being autogamous, this result was expected for *C. arabica*, since most of the individuals that compose the population evaluated in this study were derived from contrasting crosses.

#### Study of population structure

Similar to the dendrogram analysis, population structure formed two groups. The first group clustered all the accessions of the germplasm bank, which evidenced the greater dissimilarity of these accessions with the HdT parents in relation to the Catuaí parents. All BCs progenies were clustered in the first group. These results can be explained by the fact that BCs progenies present greater genetic similarity with the parents of the Catuaí group. Of the three HdT parents and of the five  $F_1$  hybrids evaluated, only the parent UFV 530 (F) and the  $F_1$  hybrid H513-5 (K) were clustered in the first group. This information is of great relevance to breeding programs, since the parent HdT UFV 530 carries the  $S_{H5}$ ,  $S_{H6}$ ,  $S_{H7}$ ,  $S_{H8}$ ,  $S_{H9}$ , and  $S_{H?}$  genes, which confer durable resistance to rust (Bettencourt and Rodrigues-Júnior 1988), and is genetically closer to parents of the Catuaí group than the other HdT parents. Accordingly, in a study on genetic variability and kinships of *Coffea* species using molecular markers, high similarity was observed between CIFC 832/1 (derived from the same plant as HdT UFV 530 [F]) and *C. arabica*, suggesting that, after initial interspecific hybridization, several spontaneous backcrosses occurred with *C. arabica* to generate Híbrido de Timor (Lashermes et al. 1993). H513-5 (K) was the only  $F_1$  hybrid clustered in group I, and this was expected, since this genotype comes from the cross between the parents UFV 2148-57 (E) and UFV 530 (F). All individuals of progeny 9 were also clustered in the first group, and were genetically clustered between the parent UFV 2148-57 (E) and the hybrid H513-5 (K). This can be explained by the genealogy of this progeny, obtained from

backcross between the hybrid H513-5 (K) and the parent of the Catuaí group UFV 2148-57 (E).

The other HdT parents, the other F<sub>1</sub> hybrids, all BCr plants, except BCr 3-2, and about 44% of the plants in F<sub>2</sub> generation were clustered in group II. These results can be explained by the introgression of *C. canephora* in *C. arabica* and prove the genetic diversity between these two germplasm, Catuaí and HdT. For being genetically divergent, these results confirm the capacity of the HdT germplasm to amplify the genetic base of *C. arabica* (Setotaw et al. 2010; Setotaw et al. 2013).

As observed in this work, a significant genetic divergence between the parents HdT UFV 530 (F) and UFV 445-46 (B) was reported in a study with RAPD, SSR and AFLP molecular markers (Setotaw et al. 2010). These authors also verified greater genetic similarity between HdT UFV 445-46 (B) and UFV 440-22 in relation to HdT UFV 530 (F). These results corroborate those observed in the present study, since HdT UFV 440-22 and UFV 440-10 (D) are derived from HdT ERU 209-6.

This result evidences the capacity of the molecular markers and of this type of genetic-statistical analysis to identify genotypes with greater recovery of the recurrent parent's genome in backcross programs. In addition, molecular markers provide more accurate and less subjective genetic information than that generated from phenotypic data, which is useful in breeding programs.

#### Genetic distance between pairs of genotypes

Genetic diversity analysis showed that genotype CIFC 134/4 is present in the greatest genetic distances detected in this study. This genotype is an accession that belongs to a germplasm bank that has not yet been genetically improved, but is important genotype for the breeding program, since it carries genes of interest that can be introgressed in commercial

cultivars. The genotype 1-5, belonging to the generation of resistant backcross (progeny 1), was involved in four of the 10 greatest genetic distances in relation to the evaluated genotypes. By the analysis of all genotypes of this progeny, the other individuals presented small genetic distance between them, whereas the mean genetic distance between genotype 1-5 and the other genotypes was high. This indicates possible mixture of pollen, during the crosses, or mixture of seedlings, during planting. By the analyses of the phenotypic characteristics of the individuals of this progeny, all the individuals presented red fruits, except for the genotype 1-5, which presented yellow fruits.

The smallest genetic distances between the pairs of genotypes under study were observed between F<sub>1</sub> hybrids H419-1 (G), H419-10 (H), H514-7 (I), and H514-8-1 (J). F<sub>1</sub> hybrids H419-1 (G) and H419-10 (H) were originated from the cross between UFV 2143-236 (A) and UFV 445-46 (B), and the hybrids H514-7 (I) and H514-8-1 (J) were originated from the cross between the parents UFV 2154-344 (C) and UFV 440-10 (D). Genetic similarity between hybrids H419-1 (G) and H419-10 (H), and between hybrids H514-7 (I) and H514-8 (J) was expected for they are full-sibs. However, similarity was observed between the four hybrids, including non-siblings. This result suggests that the parents of the crosses analyzed in this study, both the parents Catuaí (A and C) and the parents Híbrido de Timor (B and D), are genetically close. HdT UFV 445-46 (B) and UFV 440-10 (D) correspond to accessions introduced by seeds from genotypes ERU 209/15 and ERU 209/6, respectively, which originated from genotype CIFIC 2570 (Pereira et al. 2008). These data demonstrate the narrow genetic base of the breeding populations.

By the analysis of the genetic distances between the parents of all the studied populations, parents of the Catuaí group presented smaller distances in relation to HdT parents. In addition, among the Híbrido de Timor genotypes, UFV 530 (F) is the most divergent. These results demonstrate the importance of using HdT parents, especially the

most divergent, in the development of new cultivars, in order to increase the genetic distance between them. In addition, accessions of HdT coffee plants carry genes of resistance to different diseases and pests, being the main source of resistance used in breeding programs around the world. HdTs have been used in the development of rust-resistant varieties, such as Catimor and Sarchimor, which present small size, high yield, and good cup quality. Despite being genetically similar, parents of the Catuaí group stand out for the high mean yield for they have great vegetative vigor and small size (Pereira et al. 2010a). Varieties of this coffee group are flexible in relation to climatic variations, and are well-adapted to several producing regions.

The parent UFV 530 (F) was the most genetically divergent among the HdT parents; however, it was the most genetically similar to the parents of the Catuaí group. These results demonstrate the importance of using this genotype in crosses in the development of new arabica cultivars, since it carries resistance genes and is more related to the genetically improved cultivars.

By the analysis of the genetic diversity among the individuals of the same generation (F<sub>1</sub>, F<sub>2</sub>, BC<sub>r</sub> and BC<sub>s</sub>), the greatest genetic distances among genotypes were observed in the F<sub>2</sub> generation. The smallest genetic distances were observed between genotypes in F<sub>1</sub> generation. However, the values of the mean genetic distances between the four generations (F<sub>1</sub>, F<sub>2</sub>, BC<sub>r</sub> and BC<sub>s</sub>) are very close and much higher than the mean genetic distance of the parents of the Catuaí group.

An increase in the genetic distance between the individuals originated from the crosses was observed when using parents that carry resistance genes. These results confirm the importance of the introduction of HdT genotypes (which carry resistance genes) to the crosses for the development of new cultivars. Setotaw et al. (2013) also observed an increase in genetic diversity after the introduction of HdT to *C. arabica* breeding. The authors

demonstrated that rust-resistant coffee cultivars, which are basically derived from HdT, are more genetically distant than rust-susceptible coffee cultivars, and contributed to increase the genetic base of the improved Brazilian varieties.

#### Clustering Analysis

The high efficiency of SNP markers in studies on genetic diversity was observed by the analysis of the dendrogram. In spite of being highly endogamic, all the genotypes of the study population were discriminated. This indicates that these markers are powerful tools to discriminate accessions in germplasm banks, avoiding redundancy in the collections; and to discriminate cultivars, supporting the morphological descriptors (Sousa et al. 2017).

Two main groups were formed in the dendrogram. In general, rust-susceptible parents were clustered in the first group, and rust-resistant parents were clustered in the second group. The susceptible backcross progenies were also clustered in the first group. The formation of these two large groups can be explained by the fact that the rust-susceptible parents was Catuaí Amarelo, which is the most commercially planted cultivar in Brazil, and corresponds to related genotypes. The resistant parents are accessions of Híbridos de Timor; therefore, they contain introgression of *C. canephora* genome (Lashermes et al. 2000; Bertrand et al. 2003). Híbrido de Timor is a natural hybrid between *C. arabica* and *C. canephora*, in which probably an unreduced gamete of *C. canephora* was combined with another normal gamete of *C. arabica* (Bettencourt 1973). HdT presents a phenotype similar to that of *C. arabica*, it is autogamous and tetraploid with  $2n = 44$  chromosomes, being easily crossed with *C. arabica* plants. The use of HdTs in the crosses results in increased genetic diversity (Setotaw et al. 2013) for they carry several genes of *C. canephora*.

The  $F_1$  hybrids were closely clustered in group I.b, except for the hybrid H513-5 (K). The H513-5 and its progenies, 9 (BCs) and 13 ( $F_2$ ), were separated into a subgroup (I.a).

This greater diversity of H513-5 in relation to other hybrids obtained from crosses between arabica coffee and HdT was also observed by Fontes et al. (2002). The authors analyzed 29 hybrids of arabica coffee originated from crosses with different accessions of HdT, and observed that the hybrid H513-5 (K) did not cluster with any of the other evaluated genotypes. In addition, the hybrid H513-5 was resistant to rust and presented high yield capacity (Fontes et al. 2002). This F<sub>1</sub> hybrid is the result of the cross between Catuaí Amarelo IAC 64 (UFV 2148-57 [E]) and HdT UFV 530 (F). The dendrogram, in agreement with the Bayesian analysis, demonstrated that parent UFV 530 (F) is the genetically closest HdT to the parent of the Catuaí group. UFV 530 (F) has been introduced from the Coffee Rust Research Center, Portugal, to the UFV/Epamig germplasm bank by cloning the plant CIFC 832/2. This genotype is considered as one of the main sources of resistance to rust, since it contains at least six resistance genes (Eskes 1989), and its resistance has not been supplanted yet. The other HdT parents of the studied progenies were introduced to the germplasm bank by using seeds of the same plant, CIFC 2570, justifying the genetic proximity of these HdTs and their F<sub>1</sub> hybrids in the dendrogram.

Three accessions belonging to the germplasm bank, CIFC 128/2, CIFC 635/3, and CIFC 33/1, were also clustered in group I.a. The accessions CIFC 128/2 and CIFC 635/3 correspond to coffee plants introduced from Ethiopia and do not contain introgression of *C. canephora* into the genome, which explains why they were clustered in the same group of rust-susceptible parents. The last accession, CIFC 33/1, carries a resistance gene from another coffee species, *C. liberica* (Bettencourt and Rodrigues-Júnior 1988). The introgression of the genome of another species justifies the distance between these genotypes and the others. The greater genetic divergence and the presence of rust resistance gene different from the other coffee plants demonstrate the importance of including CIFC 33/1 in future crosses in breeding programs. The other accession of the germplasm bank, CIFC

134/4, was not clustered. This accession is not related to the parents used in this work; thus, greater genetic divergence between CIFC 134/4 and the population evaluated is expected.

#### Principal Coordinate Analysis (PCoA)

By the graphic dispersion of the scores obtained by PCoA, it was possible to observe the structuring of progenies based on the genetic distance between them. Similar to the cluster analysis, PCoA demonstrated the genetic distinction between the parent HdT UFV 530 and the F<sub>1</sub> hybrids evaluated in this study. The parents of the Catuaí group were clustered on the opposite side to that where HdT parents were clustered; F<sub>1</sub> hybrids were plotted between their respective parents; the progenies BCs and BCr were clustered close to their respective recurrent parents. These results are in agreement with other studies and can be explained by the kinship degree between the evaluated progenies.

In conclusion, the methodology of detection of the SNP molecular markers by constructing polymorphic probes was efficient to identify expressive number of SNP distributed throughout the *C. arabica* genome. The use of these SNP markers allowed the accurate discrimination of all the genotypes analyzed in the study, and even detected genetic mixtures among full-sib individuals, which are highly inbred, and therefore recommended for backcross. Analyses of the diversity and genetic structure of the population allowed evaluating the contribution of the HdT parents in the amplification of the genetic base of coffee plants. The analyses were precise in the division of the individuals into subpopulations according to the degree of introgression of *C. canephora* and its consequence in the obtainment of improved varieties was discussed. In addition, they enabled the identification of genetically different coffee plants to be introduced to breeding programs, as well as the molecular analysis of the parents that are already used.

For being numerous, polymorphic and widely distributed throughout the genome of the species, this set of markers has proved to be a useful auxiliary tool for breeding, and it

can also be used for fingerprinting of accessions of germplasm banks and cultivars, for and genomic selection and association.

**Conflict of interest**

The authors declare no conflict of interest.

**Data archiving statement**

The authors have not submitted biological data to any of the public databases.

**Table 1.** Rust resistant and susceptible *Coffea arabica* parents

<b>Code</b>	<b>Genotype</b>	<b>Description</b>	<b>Reaction to rust</b>
A	UFV 2143-236	Catuaí amarelo IAC 30	Susceptible
B	UFV 445-46	Híbrido de Timor CIFIC 4193C	Resistant
C	UFV 2154-344	Catuaí amarelo IAC 86	Susceptible
D	UFV 440-10	Híbrido de Timor CIFIC 4192B	Resistant
E	UFV 2148-57	Catuaí amarelo IAC 64	Susceptible
F	UFV 530	Híbrido de Timor CIFIC 832/2	Resistant

**Table 2.** Coffea arabica hybrid genotypes and their genealogy

Code	Hybrid	Generation	Genealogy		Reaction to rust
			Parent 1	Parent 2	
G	H419-1	F <sub>1</sub>	UFV 2143-235	UFV 445-46	Resistant
H	H419-10	F <sub>1</sub>	UFV 2143-235	UFV 445-46	Resistant
I	H514-7	F <sub>1</sub>	UFV 2154-344	UFV 440-10	Resistant
J	H514-8	F <sub>1</sub>	UFV 2154-344	UFV 440-10	Resistant
K	H513-5	F <sub>1</sub>	UFV 2148-57	UFV 530	Resistant

F<sub>1</sub> = Hybrid obtained by the cross between rust-resistant and rust-susceptible genotypes.

**Table 3.** Genotypes selected from the 13 *Coffea arabica* progenies and their genealogy

Progenies	Individuals	Genealogy	
		Parent 1	Parent 2
BCr 1	1, 2, 3, 4 and 5	H 419-1 c-17	UFV 445-46
BCs 2	1, 2, 3, 4 and 5	H 419-1 c-17	UFV 2143-235
BCr 3	1, 2, 3 and 4	H 514-8 c-387	UFV 440-10
BCs 4	1, 2, 3 and 4	H 514-8 c-387	UFV 2154-344
BCr 5	1, 2, 3 and 4	H 514-7 c-364	UFV 440-10
BCs 6	1, 2, 3 and 4	H 514-7 c-364	UFV 2154-344
BCr 7	1, 2, 3 and 4	H 419-10 c-214	UFV 445-46
BCs 8	1, 2, 3 and 4	H 419-10 c-214	UFV 2143-235
BCs 9	1, 2, 3, 4 and 5	UFV 2148-57	H513-5 c-14
F <sub>2</sub> 10	1, 2, 3 and 4	H 514-8 c-387	H 514-8 c-387
F <sub>2</sub> 11	1, 2, 3 and 4	H 514-7 c-364	H 514-7 c-364
F <sub>2</sub> 12	1, 2, 3, 4 and 5	H 419-10 c-214	H 419-10 c-214
F <sub>2</sub> 13	1, 2, 3, 4 and 5	H513-5 c-14	H513-5 c-14

BCr = Backcross of F<sub>1</sub> hybrid with recurrent resistant parent; BCs = Backcross of F<sub>1</sub> hybrid with recurrent susceptible parent; F<sub>2</sub> = generation obtained by the selfing of F<sub>1</sub> hybrids

**Table 4.** Coffea arabica accessions belonging to the germplasm bank of UFV/EPAMIG

<b>Genotypes<sup>1</sup></b>	<b>Description</b>	<b>Resistance genes</b>	<b>Reaction to rust<sup>2</sup></b>
CIFC 33/1	S288-23	S <sub>H</sub> 3 e S <sub>H</sub> 5	Resistant
CIFC 128/2	Dilla & Alghe	S <sub>H</sub> 1	Susceptible
CIFC 134/4	S 12 Kaffa	S <sub>H</sub> 1 e S <sub>H</sub> 4	Susceptible
CIFC 635/3	S 12 Kaffa	S <sub>H</sub> 1, S <sub>H</sub> 4 e S <sub>H</sub> 5	Susceptible

<sup>1</sup>Coffee differential clones of Hemileia vastatrix races; <sup>2</sup>Reação de resistência à ferrugem do cafeeiro em relação as raças de Hemileia vastatrix predominantes nas áreas agrícolas comerciais (Zambolim 2016).

**Table 5.** Number of SNP molecular markers by chromosome after filter 3, genotypic and allelic mean frequencies, and mean PIC of the SNP on each chromosome

Chromosome	N. SNP	Genotypic Frequency			Allelic Frequency		Mean PIC
		A <sup>1</sup> A <sup>1</sup>	A <sup>1</sup> A <sup>2</sup>	A <sup>2</sup> A <sup>2</sup>	A <sup>1</sup> (p)	A <sup>2</sup> (q)	
UNIGENES	238	0.2214	0.7529	0.0257	0.5978	0.4022	0.3458
0 <sup>1</sup>	2078	0.1809	0.7781	0.0410	0.5700	0.4300	0.3515
1	1566	0.1797	0.7791	0.0412	0.5692	0.4308	0.3526
2	1135	0.1794	0.7984	0.0222	0.5786	0.4214	0.3511
3	546	0.1795	0.8035	0.0170	0.5812	0.4188	0.3522
4	528	0.1803	0.7947	0.0251	0.5776	0.4224	0.3494
5	515	0.1664	0.8006	0.0331	0.5666	0.4334	0.3522
6	1037	0.1983	0.7347	0.0670	0.5656	0.4344	0.3447
7	1057	0.1847	0.7340	0.0812	0.5518	0.4482	0.3484
8	515	0.1730	0.7978	0.0292	0.5719	0.4281	0.3513
9	638	0.1812	0.7526	0.0663	0.5574	0.4426	0.3546
10	478	0.1771	0.8037	0.0192	0.5789	0.4211	0.3520
11	856	0.2064	0.7389	0.0547	0.5759	0.4241	0.3476
<b>Overall mean</b>	860.5385	0.1852	0.7745	0.0402	0.5725	0.4275	0.3503

<sup>1</sup>Chromosome "0" is just a pool of non-ordered sequence scaffolds (Denoeud et al. 2014).

**Table 6.** The 10 greatest and the 10 smallest genetic distances between the pairs of Arabica coffee genotypes obtained with the SNP analysis

Genotype i	Genotype i'	Greatest distances i i'	Genotype i	Genotype i'	Smallest distances i i'
CIFC 134/4	UFV 440-10 (D)	0.2542	H514-7 (I)	H419-10 (H)	0.0673
CIFC 134/4	BCr 1-5 <sup>1</sup>	0.2529	H419-1 (G)	H419-10 (H)	0.0742
CIFC 33/1	BCr 1-5	0.2477	H514-7 (I)	H419-1 (G)	0.0796
CIFC 134/4	BCr 7-2	0.2469	H514-7 (I)	BCs 2-5	0.0813
CIFC 134/4	BCr 7-3	0.2424	BCs 9-4	BCs 9-1	0.0817
CIFC 134/4	UFV 445-46 (B)	0.2382	H419-10 (H)	BCs 2-5	0.0818
CIFC 134/4	BCr 1-1	0.2369	F <sub>2</sub> 12-2	BCs 4-3	0.0834
UFV 2148-57 (E)	BCr 1-5	0.2369	H419-10 (H)	BCs 2-4	0.0834
CIFC 134/4	BCr 5-2	0.2358	H514-8 (J)	H514-7 (I)	0.0835
UFV 2154-344 (C)	BCr 1-5	0.2344	BCr 1-3	BCr 3-1	0.0835

<sup>1</sup>The first number, from 1 to 12, corresponds to the progeny, and the second number, from 1 to 5, corresponds to the plant number within the progeny, as listed in Table 3.

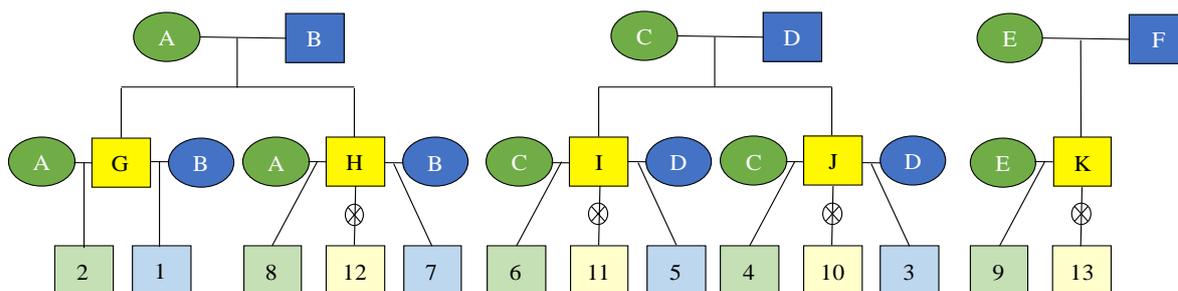
**Table 7.** Genetic distances between pairs of parents Catuaí (A, C and E) and Híbrido de Timor (B, D and F)

Distance between ii'	UFV 2154-344 (C)	UFV 2148-57 (E)	UFV 445-46 (B)	UFV 440-10 (D)	UFV 530 (F)
UFV 2143-236 (A)	0.0836	0.0940	0.2062	0.2234	0.1545
UFV 2154-344 (C)		0.0955	0.2112	0.2223	0.1616
UFV 2148-57 (E)			0.2161	0.2336	0.1648
UFV 445-46 (B)				0.1001	0.1592
UFV 440-10 (D)					0.1732

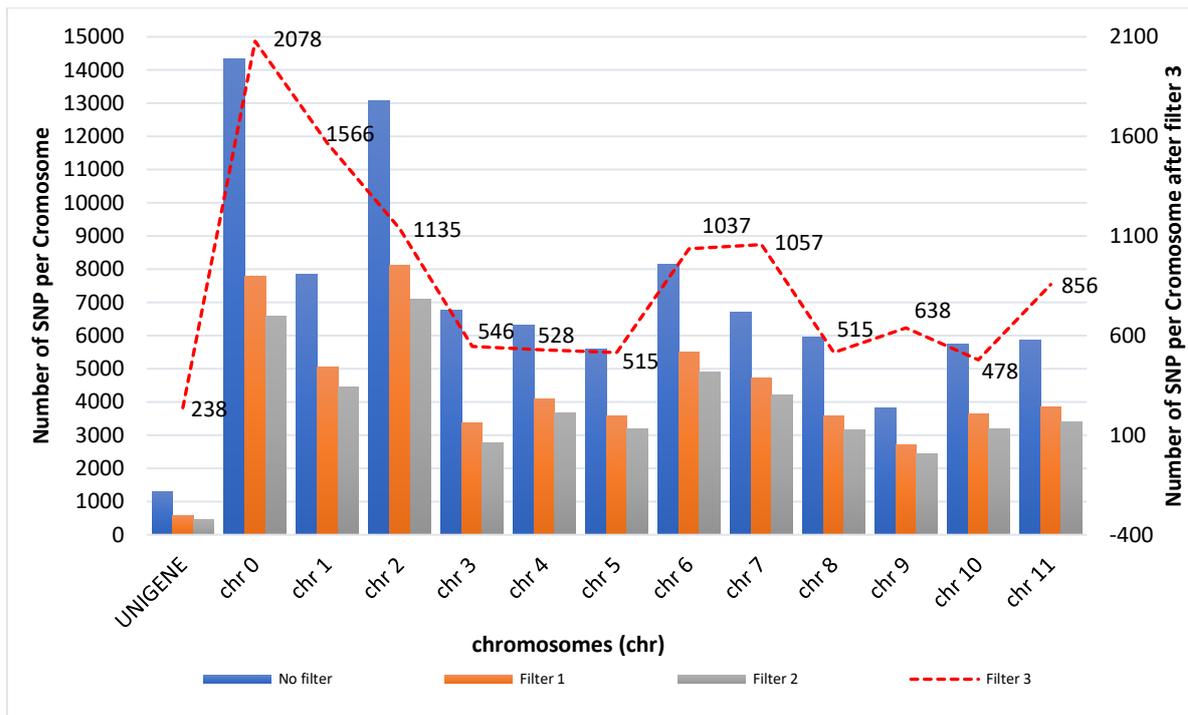
**Table 8.** The greatest and the smallest genetic distances between pairs of individuals within each group

Group	Genetic distances between ii'	
	Greatest	Smallest
F <sub>1</sub> x F <sub>1</sub>	0.1345	0.0673
F <sub>2</sub> x F <sub>2</sub>	0.1886	0.0896
BCr x BCr	0.1878	0.0835
BCs x BCs	0.1809	0.0817

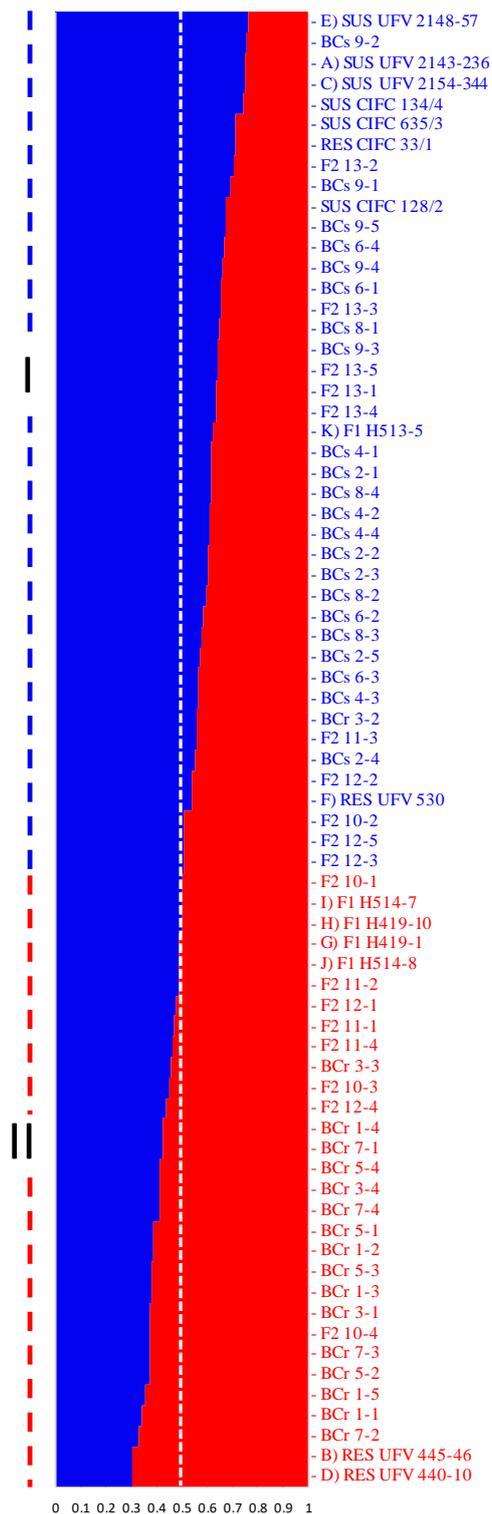
F<sub>1</sub> = generation resulting from crosses between contrasting genotypes for resistance to coffee rust; F<sub>2</sub> = generation obtained by the selfing of F<sub>1</sub> hybrids; BCr = Backcross with recurrent resistant parent; BCs = Backcross with recurrent susceptible parent.



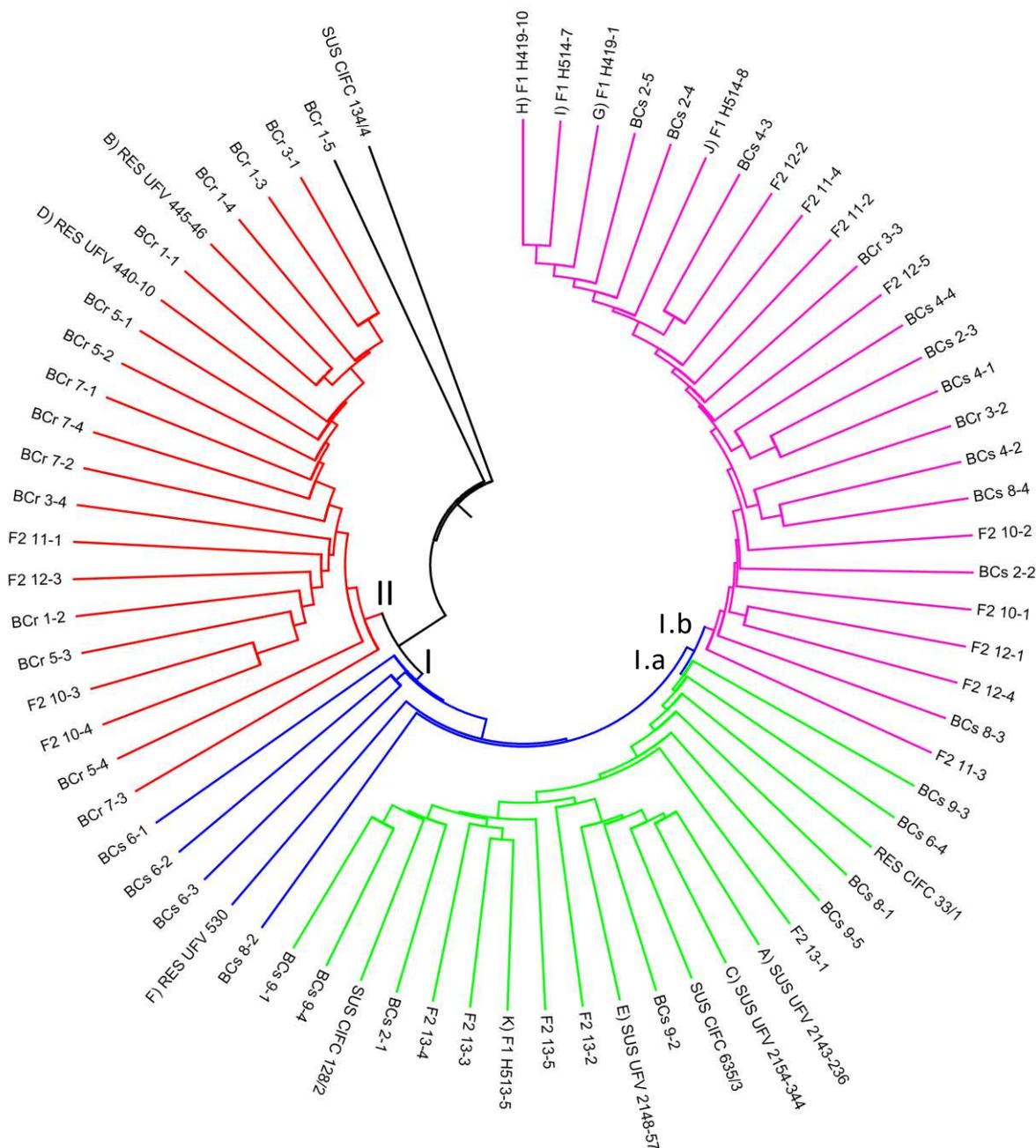
**Fig. 1** Heredogram of the parents, hybrids and of the 13 progenies that compose the *Coffea arabica* breeding program analyzed in this study; Tables 1, 2 and 3 present the relation of the respective genotypes corresponding to each letter and number;  $\otimes$  = Selfing generation.



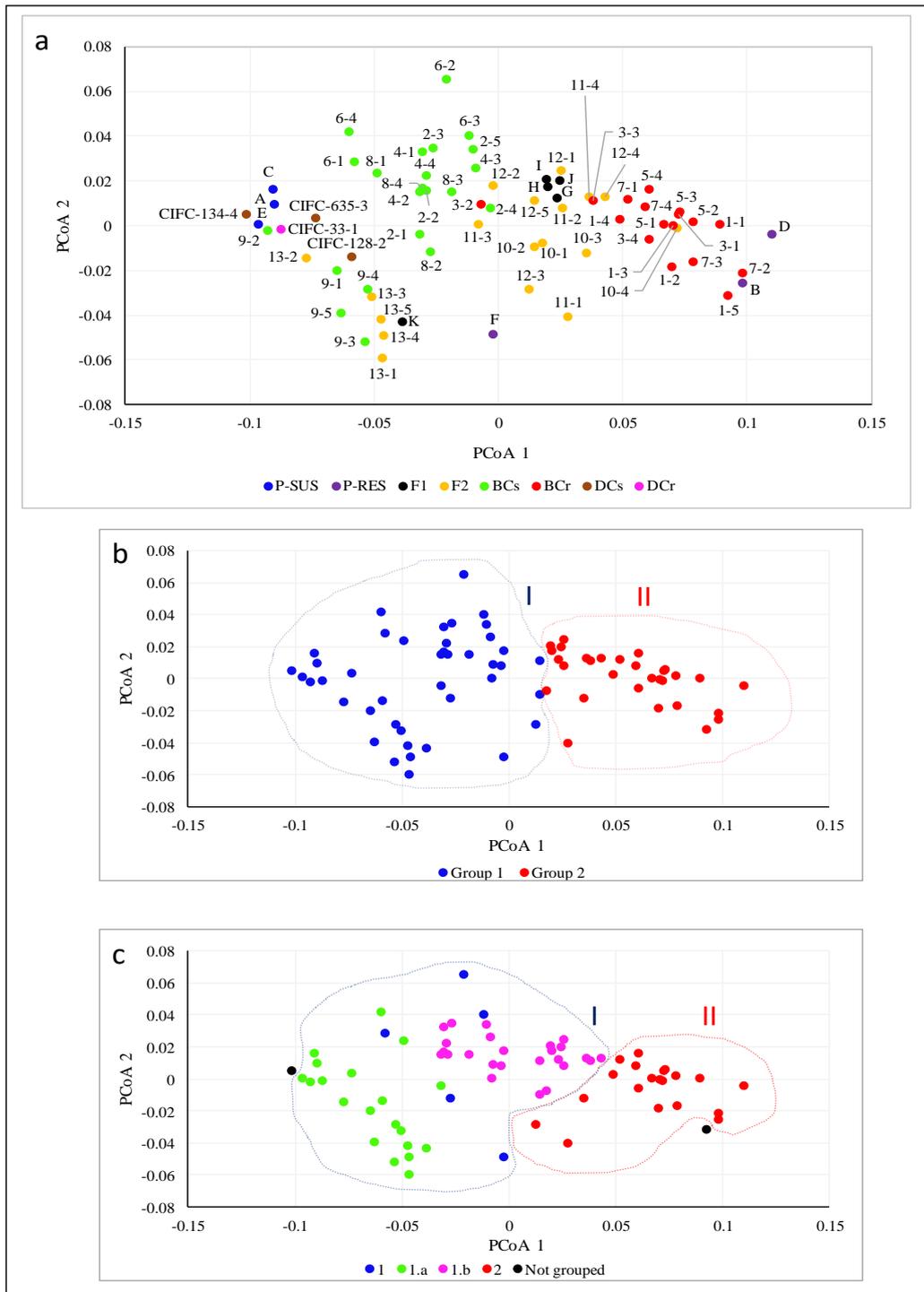
**Fig. 2** SNP molecular markers distributed throughout the UNIGENES from the EST sequences of *Coffea arabica* and of the 11 *Coffea canephora* chromosomes and chromosome "0" of *Coffea canephora*. Chromosome "0" is just a pool of non-ordered sequence scaffolds (Denoeud et al. 2014).



**Fig. 3** Bar graphs of the InStruct software used to determine the population structure of the 72 *Coffea arabica* genotypes, showing the formation of two groups ( $k = 2$ ); Tables 1 and 2 list the genotypes corresponding to each letter; SUS = Coffee rust susceptible genotype; RES = Coffee rust resistant genotype; BCr = Backcross of F<sub>1</sub> hybrid with resistant recurrent parent; BCs = Backcross of F<sub>1</sub> hybrid with recurrent susceptible parent; F<sub>2</sub> = generation obtained by the selfing of F<sub>1</sub> hybrids



**Fig. 4** Dendrogram obtained by the UPGMA technique based on the dissimilarity matrix of the arithmetic complement of the weighted index from 72 *Coffea arabica* genotypes; The list of the genotypes corresponding to each letter, from A to K, are presented in Tables 1 and 2; SUS = coffee rust susceptible genotype ; RES = coffee rust resistant genotype; BCr = Backcross of F<sub>1</sub> hybrid with recurrent resistant parent; BCs = Backcross of F<sub>1</sub> hybrid with recurrent susceptible parent; F<sub>2</sub> = generation obtained by the selfing of F<sub>1</sub> hybrids.



**Fig. 5** Principal Coordinates Analysis (PCoA) of the 72 *Coffea arabica* genotypes; a) Groups formed according to type of generation; b) Groups formed according to analysis in InStruct software; c) Groups formed according to dendrogram analysis; P-SUS = Parents of the Catuaí group; P-RES = Parents of Híbrido de Timor; F<sub>1</sub> = Hybrid obtained by the cross between rust-resistant and rust-susceptible genotypes; F<sub>2</sub> = generation obtained by the selfing of F<sub>1</sub> hybrids; BCr = Backcross of F<sub>1</sub> hybrid with recurrent resistant parent; BCs = Backcross of F<sub>1</sub> hybrid with recurrent susceptible parent; DCr = Resistant Coffee differential clones of *Hemileia vastatrix* races; DCs = Susceptible Coffee differential clones of *Hemileia vastatrix* races; The list of the genotypes corresponding to each letter, from A to K, are presented in Tables 1 and 2.

## REFERENCES

- Aerts R, Berecha G, Gijbels P, et al (2013) Genetic variation and risks of introgression in the wild *Coffea arabica* gene pool in south-western Ethiopian montane rainforests. *Evolutionary Applications* 6:243–252. doi: 10.1111/j.1752-4571.2012.00285.x
- Alkimim ER, Caixeta ET, Sousa TV, et al (2017) Marker-assisted selection provides arabica coffee with genes from other *Coffea* species targeting on multiple resistance to rust and coffee berry disease. *Molecular Breeding* 37:6. doi: 10.1007/s11032-016-0609-1
- Azevedo Peixoto L de, Laviola BG, Alves AA, et al (2017) Breeding *Jatropha curcas* by genomic selection: A pilot assessment of the accuracy of predictive models. *PLOS ONE* 12:e0173368. doi: 10.1371/journal.pone.0173368
- Bertrand B, Anthony F, Lashermes P (2001) Breeding for resistance to *Meloidogyne exigua* in *Coffea arabica* by introgression of resistance genes of *Coffea canephora*. *Plant Pathology* 50:637–643. doi: 10.1046/j.1365-3059.2001.00597.x
- Bertrand B, Guyot B, Anthony F, Lashermes P (2003) Impact of the *Coffea canephora* gene introgression on beverage quality of *C. arabica*. *TAG Theoretical and Applied Genetics* 107:387–394. doi: 10.1007/s00122-003-1203-6
- Bettencourt A (1973) Considerações gerais sobre o híbrido de Timor: Origem e possibilidades de cultivo, 23rd edn. Instituto Agrônômico, Campinas
- Bettencourt A, Rodrigues-Júnior C (1988) Principles and practice of coffee breeding for resistance to rust and other diseases. In: Clarke RJ, Macrae R (eds) *Coffee Agronomy*. Elsevier Applied Science Publishers LTD, London, pp 199–234
- Bhering LL (2017) Rbio: A tool for biometric and statistical analysis using the R platform. *Crop Breeding and Applied Biotechnology* 17:187–190. doi: 10.1590/1984-70332017v17n2s29
- Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map

- in man using restriction fragment length polymorphisms. *American journal of human genetics* 32:314–31.
- Carvalho A, Krug CA (1949) Agentes de polinização da flor do cafeeiro (*Coffea arabica* L.). *Bragantia*. doi: 10.1590/S0006-87051949000100002
- Ceccarelli S (2015) Efficiency of Plant Breeding. *Crop Science* 55:87. doi: 10.2135/cropsci2014.02.0158
- Clarindo WR, Carvalho CR (2008) First *Coffea arabica* karyogram showing that this species is a true allotetraploid. *Plant Systematics and Evolution* 274:237–241. doi: 10.1007/s00606-008-0050-y
- Combes MC, Andrzejewski S, Anthony F, et al (2000) Characterization of microsatellite loci in *Coffea arabica* and related coffee species. *Molecular Ecology* 9:1178–1180. doi: 10.1046/j.1365-294x.2000.00954-5.x
- Crossa J, Campos G d. I., Perez P, et al (2010) Prediction of Genetic Values of Quantitative Traits in Plant Breeding Using Pedigree and Molecular Markers. *Genetics* 186:713–724. doi: 10.1534/genetics.110.118521
- Cruz CD (2013) GENES - a software package for analysis in experimental statistics and quantitative genetics. *Acta Scientiarum Agronomy* 35:271–276. doi: 10.4025/actasciagron.v35i3.21251
- Cubry P, Musoli P, Legnate H, et al (2008) Diversity in coffee assessed with SSR markers: Structure of the genus *Coffea* and perspectives for breeding. *Genome* 51:50–63. doi: 10.1139/G07-096
- Danecek P, Auton A, Abecasis G, et al (2011) The variant call format and VCFtools. *Bioinformatics* 27:2156–2158. doi: 10.1093/bioinformatics/btr330
- Davis AP (2010) Six species of *Psilanthus* transferred to *Coffea* (Coffeeae, Rubiaceae). *Phytotaxa* 10:41–45.

- Davis AP (2011) *Psilanthus mannii*, the type species of *Psilanthus*, transferred to *Coffea*. *Nordic Journal of Botany* 29:471–472. doi: 10.1111/j.1756-1051.2011.01113.x
- Davis AP, Govaerts R, Bridson DM, Stoffelen P (2006) An annotated taxonomic conspectus of the genus *Coffea* (Rubiaceae). *Botanical Journal of the Linnean Society* 152:465–512. doi: 10.1111/j.1095-8339.2006.00584.x
- Denoeud F, Carretero-Paulet L, Dereeper A, et al (2014) The coffee genome provides insight into the convergent evolution of caffeine biosynthesis. *Science* (New York, NY) 345:1181–1184. doi: 10.1126/science.1255274
- Diniz LEC, Sakiyama NS, Lashermes P, et al (2005) Analysis of AFLP markers associated to the Mex-1 resistance locus in Icatu progenies. *Crop Breeding and Applied Biotechnology* 5:387–393. doi: 10.12702/1984-7033.v05n04a03
- Diola V, de Brito GG, Caixeta ET, et al (2011) High-density genetic mapping for coffee leaf rust resistance. *Tree Genetics & Genomes* 7:1199–1208. doi: 10.1007/s11295-011-0406-2
- Elsik CG, Tellam RL, Worley KC (2009) The Genome Sequence of Taurine Cattle: A window to ruminant biology and evolution. *Science* (New York, NY) 324:522–528. doi: 10.1126/science.1169588
- Esques AB (1989) Resistance. In: Kushalapa AC, Esques AB (eds) *Coffee Rust: Epidemiology, Resistance and Management*, CRC Press. Boca Raton, Florida, pp 171–293
- Ferrão LF V, Caixeta ET, Pena G, et al (2015) New EST–SSR markers of *Coffea arabica*: transferability and application to studies of molecular characterization and genetic mapping. *Molecular Breeding* 35:31. doi: 10.1007/s11032-015-0247-z
- Fontes JRM, Sakiyama NS, Cardoso AA, et al (2002) Avaliação de híbridos F1 de café (*Coffea arabica* L.) e respectivos progenitores com marcadores RAPD / Evaluation of F1 coffee hybrids (*Coffea arabica* L.) and their respective progenitors with RAPD

- markers. *Ceres* 49:283–294.
- Gao H, Williamson S, Bustamante CD (2007) A Markov Chain Monte Carlo Approach for Joint Inference of Population Structure and Inbreeding Rates From Multilocus Genotype Data. *Genetics* 176:1635–1651. doi: 10.1534/genetics.107.072371
- Garcia C, Lima B, Almeida A, et al (2011) Genome wide selection for Eucalyptus improvement at international paper in Brazil. *BMC Proceedings* 5:44. doi: 10.1186/1753-6561-5-S7-P44
- Gichimu B, Gicheru E, Mamati G, Nyende A (2013) Variation and association of cup quality attributes and resistance to Coffee Berry Disease in *Coffea arabica* L. composite cultivar, Ruiru 11. *African Journal of Horticultural Science* 7:22–35.
- Gichimu BM, Gichuru EK, Mamati GE, Nyende AB (2014) Occurrence of Ck-1 gene conferring resistance to Coffee Berry Disease in *Coffea arabica* cv. Ruiru 11 and its parental genotypes. *Journal of Agricultural and Crop Research* 2:51–61.
- Hallauer AR (2011) Evolution of plant breeding. *Crop Breeding and Applied Biotechnology* 11:197–206. doi: 10.1590/S1984-70332011000300001
- Heffner EL, Sorrells ME, Jannink J-L (2009) Genomic Selection for Crop Improvement. *Crop Science* 49:1. doi: 10.2135/cropsci2008.08.0512
- Hendre P, Phanindranath R, Annapurna V, et al (2008) Development of new genomic microsatellite markers from robusta coffee (*Coffea canephora* Pierre ex A. Froehner) showing broad cross-species transferability and utility in genetic studies. *BMC Plant Biology* 8:51. doi: 10.1186/1471-2229-8-51
- Heslot N, Yang H-P, Sorrells ME, Jannink J-L (2012) Genomic Selection in Plant Breeding: A Comparison of Models. *Crop Science* 52:146. doi: 10.2135/cropsci2011.06.0297
- ICO (2017) International Coffee Organization. In: Note regarding the review of statistical data published by the ICO. <http://www.ico.org/prices/po-production.pdf>. Accessed 6

Apr 2017

- Inácio P, Lewinsohn T, Carmo RL do, Hogan DJ (2002) Ordenação multivariada na ecologia e seu uso em ciências ambientais. *Ambiente & Sociedade* 69–83. doi: 10.1590/S1414-753X2002000100005
- Kumar S, Stecher G, Tamura K (2016) MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular Biology and Evolution* 33:1870–1874. doi: 10.1093/molbev/msw054
- Lashermes P, Andrzejewski S, Bertrand B, et al (2000) Molecular analysis of introgressive breeding in coffee (*Coffea arabica* L.). *TAG Theoretical and Applied Genetics* 100:139–146. doi: 10.1007/s001220050019
- Lashermes P, Combes M-C, Robert J, et al (1999) Molecular characterisation and origin of the *Coffea arabica* L. genome. *Molecular and General Genetics MGG* 261:259–266. doi: 10.1007/s004380050965
- Lashermes P, Combes MC, Ansaldi C, et al (2011) Analysis of alien introgression in coffee tree (*Coffea arabica* L.). *Molecular Breeding* 27:223–232. doi: 10.1007/s11032-010-9424-2
- Lashermes P, Cros J, Marmey P, Charrier A (1993) Use of random amplified DNA markers to analyse genetic variability and relationships of *Coffea* species. *Genetic Resources and Crop Evolution* 40:91–99. doi: 10.1007/BF00052639
- Laurie CC, Doheny KF, Mirel DB, et al (2010) Quality control and quality assurance in genotypic data for genome-wide association studies. *Genetic Epidemiology* 34:591–602. doi: 10.1002/gepi.20516
- Maluf MP, Silvestrini M, Ruggiero LM de C, et al (2005) Genetic diversity of cultivated *Coffea arabica* inbred lines assessed by RAPD, AFLP and SSR marker systems. *Scientia Agricola* 62:366–373. doi: 10.1590/S0103-90162005000400010

- Matukumalli LK, Lawley CT, Schnabel RD, et al (2009) Development and Characterization of a High Density SNP Genotyping Assay for Cattle. *PLoS ONE* 4:e5350. doi: 10.1371/journal.pone.0005350
- Missio RF, Caixeta ET, Zambolim EM, et al (2009a) Development and validation of SSR markers for *Coffea arabica* L. *Crop Breeding and Applied Biotechnology* 9:361–371. doi: 10.12702/1984-7033.v09n04a11
- Missio RF, Caixeta ET, Zambolim EM, et al (2011) Genetic characterization of an elite coffee germplasm assessed by gSSR and EST-SSR markers. *Genetics and Molecular Research* 10:2366–2381. doi: 10.4238/2011.October.6.2
- Missio RF, Caixeta ET, Zambolim EM, et al (2009b) Assessment of EST-SSR markers for genetic analysis on coffee. *Bragantia* 68:573–581. doi: 10.1590/S0006-87052009000300003
- Moncada MDP, Tovar E, Montoya JC, et al (2016) A genetic linkage map of coffee (*Coffea arabica* L.) and QTL for yield, plant height, and bean size. *Tree Genetics & Genomes* 12:5. doi: 10.1007/s11295-015-0927-1
- Moncada P, McCouch S (2004) Simple sequence repeat diversity in diploid and tetraploid *Coffea* species. *Genome* 47:501–509. doi: 10.1139/g03-129
- Oliveira ACB, Sakiyama NS, Caixeta ET, et al (2007) Partial map of *Coffea arabica* L. and recovery of the recurrent parent in backcross progenies. *Crop Breeding and Applied Biotechnology* 7:196–203.
- Ortiz R, Lund B, Andersen SB (2003) Breeding gains and changes in morphotype of Nordic spring wheat (1901-1993) under contrasting environments. *Genetic Resources and Crop Evolution* 50:455–459. doi: 10.1023/A:1023902110224
- Pearl HM, Nagai C, Moore PH, et al (2004) Construction of a genetic map for arabica coffee. *TAG Theoretical and Applied Genetics* 108:829–835. doi: 10.1007/s00122-003-1498-

- Pereira AA, Carvalho GR, Moura WM, et al (2010a) Cultivares: Origem e suas Características. In: Reis PR, Cunha RL (eds) Café arábica do plantio à colheita, EPAMIG. Lavras, pp 167–221
- Pereira AA, Oliveira ACB, Sakiyama NS (2008) Híbrido de Timor como fonte de resistência a doenças e de qualidade da bebida do cafeeiro. In: FERNANDES LH. (ed) Manejo Fitossanitário da Cultura do Cafeeiro. Sociedade Brasileira de Fitopatologia, Brasília-DF, pp 13–24
- Pereira MC, Chalfoun SM, Carvalho GR de, Savian TV (2010b) Multivariate analysis of sensory characteristics of coffee grains (*Coffea arabica* L.) in the region of upper Paranaíba. *Acta Scientiarum Agronomy*. doi: 10.4025/actasciagron.v32i4.4283
- Pereira TB, Setotaw TA, Santos DN, et al (2016) Identification of microsatellite markers in coffee associated with resistance to *Meloidogyne exigua*. *Genetics and Molecular Research*. doi: 10.4238/gmr.15038054
- Pestana KN, Capucho AS, Caixeta ET, et al (2015) Inheritance study and linkage mapping of resistance loci to *Hemileia vastatrix* in Híbrido de Timor UFV 443-03. *Tree Genetics & Genomes* 11:72. doi: 10.1007/s11295-015-0903-9
- Poncet V, Rondeau M, Tranchant C, et al (2006) SSR mining in coffee tree EST databases: potential use of EST–SSRs as markers for the *Coffea* genus. *Molecular Genetics and Genomics* 276:436–449. doi: 10.1007/s00438-006-0153-5
- Resende M, Caixeta E, Alkimim ER, et al (2016) High-Throughput Targeted Genotyping of *Coffea arabica* and *Coffea canephora* Using Next Generation Sequencing. San Diego, CA, p 1
- Resende MFRJ, Muñoz P, Acosta JJ, et al (2012a) Accelerating the domestication of trees using genomic selection: accuracy of prediction models across ages and environments.

- New Phytologist 193:617–624. doi: 10.1111/j.1469-8137.2011.03895.x
- Resende MFRJ, Munoz P, Resende MD V., et al (2012b) Accuracy of Genomic Selection Methods in a Standard Data Set of Loblolly Pine (*Pinus taeda* L.). *Genetics* 190:1503–1510. doi: 10.1534/genetics.111.137026
- Rodgers DM, Murphy JP, Frey KJ (1983) Impact of Plant Breeding on the Grain Yield and Genetic Diversity of Spring Oats. *Crop Science* 23:737. doi: 10.2135/cropsci1983.0011183X002300040032x
- Romero G, Vásquez LM, Lashermes P, Herrera JC (2014) Identification of a major QTL for adult plant resistance to coffee leaf rust (*Hemileia vastatrix*) in the natural Timor hybrid (*Coffea arabica* x *C. canephora*). *Plant Breeding* 133:121–129. doi: 10.1111/pbr.12127
- Rovelli P, Mettulio R, Anthony F, et al (2000) Microsatellites in *Coffea arabica* L. In: Sera T, Soccol C, Pandey A, Roussos S (eds) *Coffee Biotechnology and Quality*. Springer Netherlands, Dordrecht, pp 123–133
- Setotaw TA, Caixeta ET, Pereira AA, et al (2013) Coefficient of Parentage in *Coffea arabica* L. Cultivars Grown in Brazil. *Crop Science* 53:1237–1247. doi: 10.2135/cropsci2012.09.0541
- Smith JSC, Duvick DN, Smith OS, et al (2004) Changes in Pedigree Backgrounds of Pioneer Brand Maize Hybrids Widely Grown from 1930 to 1999. *Crop Science* 44:1935. doi: 10.2135/cropsci2004.1935
- Sobreira FM, Oliveira ACB de, Pereira AA, et al (2015) Sensory quality of arabica coffee (*Coffea arabica*) genealogic groups using the sensogram and content analysis. *Australian Journal of Crop Science* 9:486–493.
- Sousa TVTV, Caixeta ETET, Alkimim ERER, et al (2017) Molecular markers useful to discriminate *Coffea arabica* cultivars with high genetic similarity. *Euphytica* 213:75. doi: 10.1007/s10681-017-1865-9

- Vieira ESN, Von Pinho ÉV de R, Carvalho MGG, et al (2010) Development of microsatellite markers for identifying Brazilian *Coffea arabica* varieties. *Genetics and Molecular Biology* 33:507–514. doi: 10.1590/S1415-47572010005000055
- Vieira LGE, Andrade AC, Colombo CA, et al (2006) Brazilian coffee genome project: an EST-based genomic resource. *Brazilian Journal of Plant Physiology* 18:95–108. doi: 10.1590/S1677-04202006000100008
- Yang H-C, Lin H-C, Kang M, et al (2011) SAQC: SNP array quality control. *BMC bioinformatics* 12:100. doi: 10.1186/1471-2105-12-100
- Zambolim L (2016) Current status and management of coffee leaf rust in Brazil. *Tropical Plant Pathology* 41:1–8. doi: 10.1007/s40858-016-0065-9

## **CAPÍTULO 2**

### **Seleção de progênies de Coffea arabica via procedimento REML/BLUP**

VIÇOSA  
MINAS GERAIS - BRASIL  
2017

## RESUMO

As principais cultivares comerciais de *Coffea arabica* são pertencentes ao grupo Catuaí, as quais são suscetíveis as principais pragas e doenças dessa cultura. Para contornar essa deficiência, a partir de progênes oriundas de cruzamentos utilizando Híbridos de Timor, novas cultivares têm sido obtidas e genes de interesse, presentes em *C. canephora*, são introgrididos em *C. arabica*. A seleção das progênes superiores deve ser criteriosa e acurada, exigindo, dessa forma, métodos seletivos eficientes. O procedimento REML/BLUP possibilita estimativas acuradas de parâmetros genéticos e, por meio destes, a seleção dos melhores indivíduos nos programas de melhoramento. Assim, o objetivo foi selecionar progênes superiores e estimar os componentes de variância, herdabilidades, correlações genéticas, índice de seleção e a diversidade genética de um conjunto de progênes oriundas de cruzamentos entre Catuaí e Híbrido de Timor para as principais características de interesse agrônômico dessa cultura. Foram avaliadas 14 características agrônômicas em 228 genótipos de cafeeiros arábica, sendo estes pertencentes a 13 progênes promissoras e a duas cultivares comerciais, Catuaí Vermelho IAC 44 e Oeiras MG 6851, utilizadas como progênes testemunhas. A estimação dos parâmetros genéticos, correlações genéticas e o índice de seleção foram realizados pelo programa Selegen-REML/BLUP. Observou-se baixa variabilidade genética entre as progênes avaliadas. Utilizou-se o software GENES para realizar as análises de agrupamento e de coordenadas principais. O dendrograma evidenciou a formação de quatro grupos principais. Pela análise de coordenadas principais foi possível observar a dispersão gráfica das progênes. Pela análise do índice de seleção, selecionando as cinco progênes com os melhores ranks-médio, o ganho de seleção foi superior a 32% em relação a população original. Progênes de cafeeiros foram selecionadas por se mostrarem superiores às cultivares comerciais utilizadas como progênes testemunhas.

**Palavras chave:** Café, parâmetros genéticos, modelos mistos, análise multivariada

## **Introdução**

O melhoramento genético possibilita a obtenção de ganhos econômicos, sociais e ambientais e, em vegetais, permitiu o desenvolvimento de diversas cultivares mais produtivas e adaptadas. Para a espécie *Coffea arabica*, as cultivares mais produtivas e, conseqüentemente, mais plantadas são pertencentes ao grupo de cultivares Catuaí (Fernandes et al. 2012), destacando-se também pelo alto vigor vegetativo e porte baixo das plantas (Pereira et al. 2010a). Contudo, essas cultivares são suscetíveis às principais doenças dessa cultura, o que evidencia a necessidade de se lançar novas cultivares que reúnam as características agrônômicas de interesse das cultivares Catuaí e que sejam superiores em relação aos estresses bióticos e abióticos. No entanto, por apresentar base genética estreita (Lashermes et al. 1999; Cubry et al. 2008; Lashermes et al. 2011; Setotaw et al. 2013; Sousa et al. 2017), a obtenção de tais cultivares de *C. arabica* é tarefa desafiadora.

Para contornar essas limitações, nos últimos anos os derivados de Híbrido de Timor (HdT), híbrido natural do cruzamento entre *C. arabica* e *C. canephora* (Lashermes et al. 1999), tem sido utilizado nos blocos de cruzamentos para obtenção de progênies por facilitarem a introgressão de alelos de interesse de *C. canephora* em *C. arabica*. Este germoplasma é portador de genes que conferem resistência durável à *Hemileia vastatrix* (Diola et al. 2011; Romero et al. 2014; Alkimim et al. 2017), *Colletotrichum kahawae* (Gichimu et al. 2013; Gichimu et al. 2014; Alkimim et al. 2017) e *Meloidogyne exigua* (Bertrand et al. 2001). Além disso, cultivares derivadas de HdT demonstraram superioridade na qualidade de bebida em relação às cultivares tradicionais de café arábica (Pereira et al. 2010b; Sobreira et al. 2015) e tem possibilitado o aumento da base genética de *C. arabica* (Setotaw et al. 2013).

Em busca de cultivares que reúnam o máximo de vantagens agronômicas, os programas de melhoramento de *C. arabica* realizam hibridações entre genitores Catuaí e HdT, obtendo progênies que são avaliadas e, as melhores, selecionadas. Esta seleção deve ser acurada, pois pode definir se o programa de melhoramento terá sucesso, e em espécies perenes, por apresentar longo período juvenil, isso é intensificado. Nessas espécies, os caracteres fenotípicos podem se expressar espaçados por longos períodos de dias, meses ou anos. Nesse sentido, é evidente a importância de utilizar métodos seletivos acurados nos programas de melhoramento para o desenvolvimento das cultivares.

Basicamente, os métodos seletivos utilizados nos programas de melhoramento genético têm sido realizados pela avaliação fenotípica dos indivíduos em um local e ano. Assim, apresentam limitações, pois as características fenotípicas podem ser altamente influenciadas pelo ambiente (Hoffmann et al. 2005; Andrew et al. 2010), sobretudo, quando estas são governadas por muitos genes e/ou apresentam herdabilidades baixas. Além disso, em plantas perenes como o café arábica que apresentam longo período juvenil e bienalidade produtiva, a interação genótipo ambiente (colheitas) torna-se fator limitante para esse tipo de seleção fenotípica (Andrew et al. 2010). Dessa forma, métodos seletivos que se baseiam na seleção de progênies avaliadas em mais de um ano e/ou local têm sido mais precisos e têm proporcionado maiores valores de acurácia (Pereira et al. 2013).

Em geral, os dados experimentais são avaliados por meio de análises de variância (ANOVA) e, em seguida, por testes de médias e/ou regressão. A ANOVA apresenta pressuposições básicas que devem ser atendidas, tais como a independência dos erros e a independência entre os efeitos de tratamentos. Essas pressuposições nem sempre são possíveis de serem atendidas (Resende 2007) sendo, portanto, limitação para esses procedimentos estatísticos. Por permitir relaxar estas suposições, o método REML (Residual or Restricted Maximum Likelihood) possibilita modelagens mais flexíveis e análise de

tratamentos correlacionados, como ocorre em progênies geneticamente aparentadas (Resende 2007).

A metodologia de modelos mistos, REML/BLUP - Residual or Restricted Maximum Likelihood/Best Linear Unbiased Prediction (Patterson and Thompson 1971; Henderson 1975) tem se mostrado vantajosa por permitir, a partir de informações fenotípicas, a predição acurada e não viesada dos valores genéticos de indivíduos. Isso permite, simultaneamente, estimar os componentes da variação genética e os valores genéticos individuais com maior precisão (Piepho et al. 2008). Esse método preditivo é recomendado, inclusive, em experimentos com dados desbalanceados e foi aplicado com sucesso em muitas espécies vegetais (Resende and Thompson 2004; Viana et al. 2011; Ferreira et al. 2012; Barbosa et al. 2012; Pereira et al. 2013; Spinelli et al. 2015; Corrêa et al. 2015).

Nesse trabalho, objetivou-se estimar, via metodologia dos modelos mistos, REML/BLUP, os componentes de variância, herdabilidades e as correlações de características agrônômicas em progênies de *C. arabica* oriundas do cruzamento entre genitores do grupo de cultivares Catuaí e Híbrido de Timor; selecionar as melhores progênies com base no índice de seleção; quantificar o ganho genético obtido pela seleção; e estudar a diversidade genética das progênies de cafeeiros arábica.

## **Material e métodos**

### **Condução experimental**

As avaliações do experimento foram realizadas em campo experimental do Departamento de Fitopatologia da Universidade Federal de Viçosa, em Viçosa, MG (20°44'25" S; 42°50'52" W), nos anos de 2014, 2015 e 2016. Foi realizado a calagem do solo e adubação de plantio conforme exigência da cultura. Os genótipos foram plantados no dia 11 de fevereiro de 2011. As plantas foram dispostas em espaçamento de 3,0 m entre fileiras

e 0,7 m entre plantas. Após o plantio, não foram aplicados nenhum método de controle fitossanitário para controle de ferrugem, antracnose e bicho mineiro.

### **Material genético**

A partir de cruzamentos entre genitores Catuaí e HdT, contrastantes em relação à ferrugem do cafeeiro, 13 progênes foram obtidas no programa de melhoramento genético de *C. arabica* da Epamig/UFV/Embrapa (Figura 1). Estas progênes são gerações de retrocruzamento resistente (RCr), retrocruzamento suscetível (RCs) e F<sub>2</sub>. Foram selecionados 15 genótipos em cada progênie. Além disso, foram avaliados, como progênes testemunhas, 16 e 17 genótipos das cultivares Catuaí IAC 44 e Oeiras MG 6851, respectivamente, para estimar os efeitos de ambiente. Assim, foram avaliados 228 genótipos individuais (Tabela 1).

### **Avaliações Fenotípicas**

Foram realizadas avaliações fenotípicas de 14 características agrônômicas (Tabela 2), sete categóricas e sete contínuas, nos 228 genótipos de *C. arabica* relacionados na Tabela 1.

As características contínuas foram avaliadas pela mensuração fenotípica conforme descrito na tabela 2. Já as características categóricas, conforme detalhado a seguir, foram avaliadas por notas. O tamanho dos frutos maduros foi avaliado com notas de 1, 2 e 3 para frutos pequenos, médios e graúdos, respectivamente. A uniformidade de maturação foi avaliada por notas de 1 a 4, sendo as notas 1, 2, 3 e 4 para uniforme, semiuniforme, semidesuniforme e desuniforme, respectivamente. O ciclo de maturação dos frutos foi avaliado com notas de 1 a 5, sendo notas 1, 2, 3, 4 e 5 para precoce, semiprecoce, médio, semitardio e tardio, respectivamente. As incidências à ferrugem do cafeeiro, à cercosporiose e ao bicho mineiro foram avaliadas com notas de 1 a 5, onde as notas 1 e 5 foram atribuídas

para genótipos sem nenhum sintoma das enfermidades e para genótipos altamente suscetíveis às mesmas, respectivamente. A característica vigor vegetativo foi avaliada por notas de 1 a 10, onde a nota 1 foi atribuída a plantas totalmente depauperada e a nota 10 para plantas consideradas de vigor vegetativo máximo.

### **Análises Estatísticas**

As análises estatísticas e biométricas foram realizadas considerando os modelos lineares mistos (procedimento REML/BLUP), por meio do software Selegen-REML/BLUP (Resende 2016), cujos componentes de variância são estimados pelo método da máxima verossimilhança restrita (REML) e os valores genotípicos preditos pela melhor predição linear não viesada (BLUP).

A estimação dos parâmetros genéticos foi realizada pela análise individual dos 14 caracteres utilizando o seguinte modelo estatístico:

$$y = Xu + Zg + Wp + Vr + Tb + Ri + e$$

Em que,

$y$  é o vetor de dados;

$u$  é o vetor referente à média geral em cada ano de avaliação;

$g$  é o vetor de efeitos de progênies (efeito aleatório);

$p$  é variância permanente entre plantas (efeito aleatório);

$r$  é variância entre tipos de populações (efeito aleatório);

$b$  é variância entre parcela (efeito aleatório);

$i$  é variância da interação progênies X anos (efeito aleatório);

$e$  é o vetor de resíduos (efeito aleatório).

As letras maiúsculas representam as matrizes de incidência para os referidos efeitos.

As estruturas de médias e variâncias associadas ao modelo são:

$$y|u, V \sim N(Xu, V); \quad g|G, \sigma_g^2 \sim N(0, G\sigma_g^2); \quad p|\sigma_p^2 \sim N(0, I\sigma_p^2); \quad r|\sigma_r^2 \sim N(0, I\sigma_r^2); \\ b|\sigma_b^2 \sim N(0, I\sigma_b^2); \quad i|\sigma_i^2 \sim N(0, I\sigma_i^2); \quad e|\sigma_e^2 \sim N(0, I\sigma_e^2).$$

As covariâncias dos efeitos são iguais à zero. Em que:

$V$  = Variância fenotípica e equivale a:

$$ZZ'\sigma_g^2 + WW'\sigma_p^2 + VV'\sigma_r^2 + TT'\sigma_b^2 + RR'\sigma_i^2 + I\sigma_e^2;$$

$G$  = Matriz de correlação genotípica entre as progênes em avaliação;

$I$  = Matriz Identidade;

$\sigma_g^2, \sigma_p^2, \sigma_r^2, \sigma_b^2, \sigma_i^2$  e  $\sigma_e^2$  são as variâncias genotípica entre progênes, de ambiente permanente, entre tipos de progênes, de parcela, da interação progênes x anos, e residual, respectivamente.

As equações de modelos mistos para o BLUP (Best Linear Unbiased Predictor) dos efeitos aleatórios equivalem a:

$$\begin{bmatrix} \hat{u} \\ \hat{g} \\ \hat{p} \\ \hat{r} \\ \hat{b} \\ \hat{i} \end{bmatrix} = \begin{bmatrix} X'X & X'Z & X'W & X'V & X'T & X'R \\ Z'X & Z'Z + G^{-1}\lambda_1 & Z'W & Z'V & Z'T & Z'R \\ W'X & W'Z & W'W + I\lambda_2 & W'V & W'T & W'R \\ V'X & V'Z & V'W & V'V + I\lambda_3 & V'T & V'R \\ T'X & T'Z & T'W & T'V & T'T + I\lambda_4 & T'R \\ R'X & R'Z & R'W & R'V & R'T & R'R + I\lambda_5 \end{bmatrix}^{-1} \begin{bmatrix} X'y \\ Z'y \\ W'y \\ V'y \\ T'y \\ R'y \end{bmatrix}$$

Em que:

$$\lambda_1 = \sigma_e^2 / \sigma_g^2 = \frac{1-h^2-p^2-r^2-b^2-i^2}{h^2};$$

$$\lambda_2 = \sigma_e^2 / \sigma_p^2 = \frac{1-h^2-p^2-r^2-b^2-i^2}{p^2};$$

$$\lambda_3 = \sigma_e^2 / \sigma_r^2 = \frac{1-h^2-p^2-r^2-b^2-i^2}{r^2};$$

$$\lambda_4 = \sigma_e^2 / \sigma_b^2 = \frac{1-h^2-p^2-r^2-b^2-i^2}{b^2};$$

$$\lambda_5 = \sigma_e^2 / \sigma_i^2 = \frac{1-h^2-p^2-r^2-b^2-i^2}{i^2};$$

$$h^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_p^2 + \sigma_r^2 + \sigma_b^2 + \sigma_i^2 + \sigma_e^2}: \text{herdabilidade no sentido amplo};$$

$$p^2 = \frac{\sigma_p^2}{\sigma_g^2 + \sigma_p^2 + \sigma_r^2 + \sigma_b^2 + \sigma_i^2 + \sigma_e^2}: \text{coeficiente de determinação do efeito de ambiente}$$

permanente;

$$r^2 = \frac{\sigma_r^2}{\sigma_g^2 + \sigma_p^2 + \sigma_r^2 + \sigma_b^2 + \sigma_i^2 + \sigma_e^2}: \text{coeficiente de determinação entre tipos de progênies};$$

$$b^2 = \frac{\sigma_b^2}{\sigma_g^2 + \sigma_p^2 + \sigma_r^2 + \sigma_b^2 + \sigma_i^2 + \sigma_e^2}: \text{correlação devida ao ambiente comum da parcela};$$

$$i^2 = \frac{\sigma_i^2}{\sigma_g^2 + \sigma_p^2 + \sigma_r^2 + \sigma_b^2 + \sigma_i^2 + \sigma_e^2}: \text{coeficiente de determinação do efeito da interação}$$

progênies X anos.

Os componentes de variância foram estimados por REML, Restricted Maximum Likelihood (Patterson and Thompson 1971), a partir de iterações nas equações de modelos mistos (Resende and Duarte 2007; Resende 2007), cujos estimadores são:

$$\sigma_g^2 = [\hat{g}'G^{-1}\hat{g} + \sigma_e^2 \text{tr}(G^{-1}C^{22})]/q;$$

$$\sigma_p^2 = [\hat{p}'\hat{p} + \sigma_e^2 \text{tr}(C^{33})]/N;$$

$$\sigma_{pr}^2 = [\hat{r}'\hat{r} + \sigma_e^2 \text{tr}(C^{44})]/m;$$

$$\sigma_b^2 = [\hat{b}'\hat{b} + \sigma_e^2 \text{tr}(C^{55})]/d;$$

$$\sigma_i^2 = [\hat{i}'\hat{i} + \sigma_e^2 \text{tr}(C^{66})]/l;$$

$$\sigma_e^2 = [y'y - \hat{u}'X'y - \hat{g}Z'y - \hat{p}W'y - \hat{r}V'y - \hat{b}T'y - \hat{i}R'y]/[N - r(X)].$$

Em que:

G: Matriz de correlação genotípica entre as progênies em avaliação;

C: Matriz inversa generalizada da matriz dos coeficientes das equações de modelos mistos;

tr: operador traço matricial;

r(X): posto da matriz X (número de colunas linearmente independentes);

N-r(X): número de graus de liberdade do erro;

N, q, m, d e l: número total de dados, de progênies, de tipos de progênies, de parcelas, e de combinações genótipo x ano, respectivamente.

O coeficiente de variação genética ( $CV_g$  %) foi calculado por meio da fórmula:

$$CV_g(\%) = \frac{\sqrt{\sigma_g^2}}{\mu} \times 100 \text{ e, o coeficiente de variação residual por: } CV_e(\%) = \frac{\sqrt{\sigma_e^2}}{\mu} \times 100.$$

### **Índice de seleção, correlações genéticas e diversidade genética**

Todas as características significativas pela análise de deviance foram consideradas no cálculo do índice de seleção, correlações genéticas e estudo de diversidade genética. As análises foram realizadas no software Selegen-REML/BLUP (Resende 2016).

Para o cálculo do índice de seleção, foi utilizado a metodologia do índice de rank (Mulamba and Mock 1978), com adaptações. Com base nesse índice, as progênies são classificadas de acordo com a média dos ranks (Resende 2007) obtidos de cada característica avaliada. O ganho de seleção também foi estimado com base no resultado desse índice.

As correlações genéticas entre as características foram estimadas por meio da seguinte expressão:  $r_g = \frac{\hat{\sigma}_{gxy}}{\sqrt{\hat{\sigma}_{gx}^2 \hat{\sigma}_{gy}^2}}$ , em que  $\hat{\sigma}_{gxy}$  é o estimador da covariância genética entre

as variáveis  $x$  e  $y$ ; e  $\hat{\sigma}_{gx}^2$  e  $\hat{\sigma}_{gy}^2$  são os estimadores da variância genética das variáveis  $x$  e  $y$ , respectivamente.

A diversidade genética foi avaliada pela distância genética euclidiana média entre os pares de progênies. Cada progênie foi composta por 15 plantas, exceto as progênies testemunhas Catuaí Vermelho IAC 44 e Oeiras MG 6851, que foram compostas por 16 e 17 plantas, respectivamente, conforme apresentado na tabela 1. As análises de agrupamento e de coordenadas principais foram executadas software GENES (Cruz 2013). O dendrograma foi construído utilizando a metodologia UPGMA (Unweighted Pair Group Method using Arithmetic averages) e os autovalores das coordenadas principais foram extraídos a partir da matriz de distância genética entre os pares de progênies.

## **Resultados e discussão**

### **Análise de deviance**

Foi realizada análise de deviance (ANADEV) para as 14 características fenotípicas remanescentes (Tabela 3). A ANADEV foi realizada considerando todos os efeitos simultaneamente (modelo completo) e de outras quatro maneiras distintas, nas quais, em cada análise, foi utilizado modelo reduzido para testar diferentes efeitos: 1) entre progênies; 2) permanente entre plantas; 3) entre parcelas; 4) da interação progênies x anos. Foi estimada a razão de verossimilhança a qual é obtida pela diferença entre as estimativas dos valores de deviance no modelo reduzido (sem o efeito testado) e o modelo completo. A significância de cada efeito foi verificada pelo teste da razão de máxima verossimilhança (LRT) por meio do teste de qui-quadrado, aos níveis de 1% e 5% de probabilidade, com 1 grau de liberdade.

Ao avaliar o efeito genético existente entre progênies pelo teste da razão máxima verossimilhança (LRT) foi verificado efeito significativo para as características produção, resistência à ferrugem, vigor, ciclo de maturação dos frutos e o comprimento do ramo plagiotrópico. Esses resultados demonstram a existência de variabilidade genética entre as

progênies avaliadas o que permite ganho de seleção para tais características. O efeito permanente entre plantas foi significativo para as características produção, resistência à ferrugem, vigor, ciclo de maturação dos frutos, tamanho dos frutos, comprimento de folha, comprimento do ramo plagiotrópico e para o diâmetro do caule. Esse efeito representa a influência do ambiente e dos efeitos genéticos não aditivos para determinado genótipo (Mrode 2014; Resende 2015). O efeito da variância entre parcelas foi significativo para as características resistência ao bicho mineiro, vigor, comprimento do ramo plagiotrópico e para o diâmetro do caule. Esse resultado demonstra que para essas características existiram variações ambientais entre parcelas o qual deveria ter sido controlado pelo delineamento estatístico. Em experimentos cujo delineamento em blocos casualizados (DBC) é utilizado, espera-se que dentro de cada bloco não exista variação ambiental significativa a qual é controlada pelo DBC. A interação genótipo x ambiente (GxE), sobretudo, em plantas perenes, pode alterar a classificação dos genótipos onde indivíduos apresentam comportamento adaptativo diferente em resposta a alteração ambiental (Resende et al. 2014b). O efeito da interação das progênies x anos foi significativo para as características resistência à ferrugem, vigor, ciclo de maturação dos frutos, uniformidade de maturação dos frutos e para o diâmetro do caule. Essa variação pode ocorrer por mudanças causadas por diferentes condições edafoclimáticas que ocorrem em diferentes anos o que pode promover alterações na expressão gênica dos indivíduos.

### **Componentes de variância e parâmetros genéticos**

As estimativas dos parâmetros genéticos das progênies avaliadas estão apresentadas na Tabela 4. Essa análise permitiu a decomposição da variação total (variância fenotípica) em frações conhecidas (variância genética entre progênies, variância permanente entre plantas, variância entre parcelas e variância da interação progênies x anos) e frações desconhecidas (variância residual). A maioria das características apresentaram estimativas

de variâncias genética entre progênies ( $V_g$ ) de baixa magnitude, conforme revelado pelos coeficientes de variação genética (CVg%) que, em sua maioria, foram menores que 10%. Esses resultados, no geral, demonstraram baixa variabilidade genética entre as progênies de cafeeiros avaliadas. Isto pode ser justificado pelo grau de parentesco existente entre os genitores dos cruzamentos que originaram as progênies e também pelo fato da maioria dessas progênies terem sido obtidas de retrocruzamentos. Além disso, essa espécie apresenta base genética estreita (Setotaw et al. 2013; Sousa et al. 2017). Utilizando o procedimento REML/BLUP na estimação dos parâmetros genéticos de 12 cultivares de *C. arabica* foi verificada baixa variabilidade genética (Resende et al. 2001).

As herdabilidades das características foram estimadas as quais apresentaram baixos valores para as características das populações em análise (Tabela 4). A maior (0,14) e a menor (0,01) estimativa de herdabilidade foram observadas nas características ciclo de maturação dos frutos e uniformidade de maturação dos frutos, respectivamente. O valor da herdabilidade de uma característica pode variar de 0 a 1. Baixos valores de herdabilidades são observados em características poligênicas, as quais são altamente influenciadas pelo ambiente. A herdabilidade é um coeficiente genético que expressa a relação entre a variância genotípica ( $V_g$ ) e a variância fenotípica ( $V_f$ ). O papel preditivo desse parâmetro o torna de grande importância em estudos genéticos.

As estimativas dos coeficientes de determinação permanente de plantas ( $c^2_{plan}$ ), parcelas ( $c^2_{parc}$ ) e da interação entre as progênies e os anos ( $c^2_{int}$ ) foram baixas (Tabela 4). Os baixos valores estimados para esses parâmetros podem ser explicados pelos baixos valores estimados para as variâncias referentes a cada parâmetro e, também, por esses serem muito inferiores aos valores estimados para variância fenotípica das características avaliadas. As estimativas de  $c^2_{plan}$  foram superiores as demais para a maioria das características avaliadas. Esses resultados demonstram a maior variação permanente existente entre as

plantas (cova de plantio). As baixas magnitudes de  $c^2_{\text{parc}}$  em todas as características avaliadas revela homogeneidade ambiental entre as parcelas.

Os valores estimados para repetibilidade ( $r$ ) variaram entre 2% a 39% (Tabela 4). As maiores estimativas de repetibilidade foram observadas para as características Vigor, comprimento do ramo plagiotrópico, produção, ferrugem, diâmetro do caule e tamanho dos frutos, os quais foram superiores a 25%. A repetibilidade é a correlação fenotípica intraclasse e, portanto, refere-se às correlações fenotípicas entre medições repetidas no mesmo indivíduo (Resende 2002). Esta pode ser classificada em alta ( $r \geq 0,60$ ), média ( $0,30 < r < 0,60$ ) e baixa ( $r \leq 0,30$ ) (Resende 2002). Características que apresentam altas estimativas de repetibilidade podem ter o número de medidas no indivíduo reduzido, otimizando e desonerando o programa de melhoramento (Cruz et al. 2004).

Os coeficientes de variação genético ( $CV_g$ ), residual ( $CV_e$ ) e relativos ( $CV_r$ ) foram estimados (Tabela 4). As estimativas de  $CV_e$  foram inferiores a 30% para 12, das 14 características avaliadas demonstrando boa precisão experimental. As estimativas de  $CV_g$  foram menores que as estimativas de  $CV_e$  para todas as características. Dessa forma, as estimativas de  $CV_r$ , razão entre o coeficiente de variação genotípica e o coeficiente de variação ambiental, foram menores que a unidade e variaram de 0,11 a 0,47.  $CV_r$  superior a unidade indica maior probabilidade de sucesso na identificação de genótipos superiores pelo fato da variação genética superar a variação ambiental para a característica em análise (Vencovsky 1987). Por outro lado, o  $CV_r$ , por não considerar o número de repetições experimentais, não informa sobre a acurácia seletiva (Resende and Duarte 2007).

As estimativas de acurácia seletiva ( $r_{gg}$ ) variaram de 0,40 a 0,85 (Tabela 4). A acurácia seletiva é a correlação entre os valores genéticos verdadeiros e os preditos. Valores maiores de acurácia seletiva resulta em maior confiança na avaliação fenotípica dos indivíduos. Valores de acurácias seletivas superiores a 0,70 são recomendados para as

diferentes fases dos programas de melhoramento (Resende 2007). Com número de repetições experimentais superior a cinco, valores de CVr inferiores à unidade também podem propiciar elevadas acurácia e precisão (Resende and Duarte 2007), como observado neste estudo.

### **Correlação genética**

As estimativas das correlações genéticas entre as 14 características fenotípicas estão apresentadas na tabela 5. Estatisticamente, a correlação pode ser definida como a dependência entre as funções de distribuição de duas ou mais variáveis aleatórias. Geneticamente, as correlações são causadas, principalmente, pela pleiotropia (propriedade de um gene determinar mais de uma característica fenotípica) e por ligações gênicas, estas causam correlação transitória, especialmente em populações originadas de cruzamentos entre linhagens divergentes (Resende 2015). As correlações fenotípicas podem ser desmembradas em correlações genéticas e correlações ambientais (Falconer 1987). O valor de correlação entre duas características pode variar de -1 a 1 onde valores acima de 0,70 podem ser considerados altos.

Os maiores valores, em módulo, de correlações genéticas foram observados entre as características largura e comprimento de folhas (0,83), vigor vegetativo e ciclo de maturação (0,80) e produção e vigor vegetativo (0,76). A alta e positiva correlação genética observada entre a largura e o comprimento de folhas nos permite obter ganhos simultâneos para essas duas características e, dessa forma, aumentar a área foliar. A área foliar é utilizada para investigar adaptação ecológica, competição entre plantas ou espécies e efeitos do manejo. Além disso, pode ser usada para determinar o índice de área foliar refletindo no potencial fotossintético da planta (Monteiro et al. 2005). A característica vigor vegetativo, por ter sido positiva e altamente correlacionada geneticamente com as características ciclo de maturação e produção, pode ser utilizada para seleção de progênes mais produtivas e tardias. A

característica vigor vegetativo apresentou maior estimativa de herdabilidade que a característica produção, apresenta maior facilidade de ser avaliada e pode ser avaliado precocemente, desde o início do desenvolvimento das mudas. Assim, por esses resultados, a seleção de plantas vigorosas vegetativamente permitirá a obtenção de progênes mais produtivas. Assim como observado nesse trabalho, Silvarolla et al. (1997) relataram altas estimativa de correlação de vigor vegetativo com produção e ciclo de maturação entre progênes derivadas de Híbrido de Timor. Correlação positiva entre o vigor vegetativo da planta e a produção também foi observada em outros estudos com populações de cafeeiro arábica (Carvalho et al. 1984; Bonomo et al. 2004; Teixeira et al. 2012). Os menores valores, em módulo, de correlações genéticas foram observados entre as características número de nós reprodutivos e ferrugem (0,00), comprimento de folha e comprimento do ramo plagiotrópico (0,01) e uniformidade de maturação e largura da folha (-0,03). Características que apresentam correlações genéticas com baixas magnitudes podem ser melhoradas simultaneamente com a utilização de índice de seleção (Resende et al. 2014a), contudo, a seleção de uma característica não implica, simultaneamente, em ganhos com seleção na outra.

Para os programas de melhoramento, considera-se três possíveis consequências da correlação genética entre características: 1) características que apresentem correlação alta e positiva: o melhorista pode praticar seleção na população de melhoramento considerando apenas a característica de maior herdabilidade e facilidade e, ou, menor custo de mensuração, e a população selecionada manifestará ganhos genéticos em ambas, reduzindo, desse modo, o número de características a serem selecionadas; 2) características não correlacionadas: a seleção de uma não afetará a outra; 3) características negativamente correlacionadas: a seleção para a melhoria de uma poderá não ser vantajosa, em virtude da redução na segunda (Pereira 2008), a não ser que o interesse nas características seja, simultaneamente, aumentar

em uma e reduzir na outra. As correlações genéticas entre caracteres são de grande importância para os programas de melhoramento, principalmente se a seleção em um deles apresenta dificuldades de medição e identificação e, ou, se uma determinada característica apresenta baixa herdabilidade (Cruz et al. 2004).

### **Diversidade genética**

A matriz de distâncias genéticas euclidiana média entre as progênes foi estimada e, a partir dessas informações foi construído um dendrograma e realizada a análise de componentes principais. Pela análise do dendrograma (Figura 2), considerando o corte subjetivo de 65% da maior distância genética estimada entre as progênes, foi observada a formação de quatro grupos principais. No primeiro grupo ficaram alocadas as duas progênes utilizadas como testemunhas e oito progênes. Essas oito progênes que foram alocadas com as progênes Catuaí e Oeiras, podem ser progênes promissoras uma vez que Catuaí e Oeiras são cultivares com bom desempenho agrônomico comprovado. Este primeiro grupo pode ser dividido em três subgrupos. No primeiro subgrupo, I.a, foram alocadas as duas progênes testemunhas e as progênes 7 e 13. No segundo subgrupo, I.b, foram alocadas as progênes 3 e 5. Essas progênes foram obtidas de dois genótipos híbridos  $F_1$  que são irmãos completos e a partir de retrocruzamentos resistentes utilizando o mesmo genitor doador, Híbrido de Timor UFV 440-10. No terceiro subgrupo, I.c, foram alocadas as progênes 4, 10, 11 e 12. As progênes 4 e 10 são progênes de meios-irmãos. As progênes 10, 11 e 12 são progênes  $F_2$ . As progênes 10 e 11 são progênes de irmãos-completos.

O segundo grupo foi composto pela progênie 2. No terceiro e quarto grupos foram alocadas duas progênes, progênes 8 e 9 e progênes 1 e 6, respectivamente. As progênes 8 e 9 foram obtidas de retrocruzamentos suscetíveis.

A dispersão gráfica dos escores da análise de coordenadas principais (PCoA) está apresentada na figura 3. Esta análise baseia-se na estatística multivariada que extrai os

autovalores a partir da matriz de similaridade ou dissimilaridade genética apresentando a vantagem de poder ser aplicada quando as relações entre as variáveis não são lineares (Inácio et al. 2002). Pela análise de PCoA, a correlação entre as distâncias genética e as distâncias gráficas foi de 0,56 e a distorção por distância foi de 19,03%. Foi possível observar que a maioria das progênies  $F_2$  e de retrocruzamentos resistentes ficaram agrupadas com as progênies testemunhas, Catuaí Vermelho IAC 44 e Oeiras MG 6851. Por outro lado, as progênies de retrocruzamentos suscetível ficaram mais dispersas. As análises de diversidade genética são úteis para os programas de melhoramento por evidenciar as distâncias genéticas entre os pares de indivíduos e, dessa forma, facilitam a seleção das progênies promissoras com bom desempenho agrônomo associado e distantes geneticamente. Essa análise permite a eliminação precoce de genótipos redundantes ou altamente similar o que permite ao melhorista concentrar seus esforços em genótipos geneticamente distintos permitindo, dessa forma, economia de tempo e recursos.

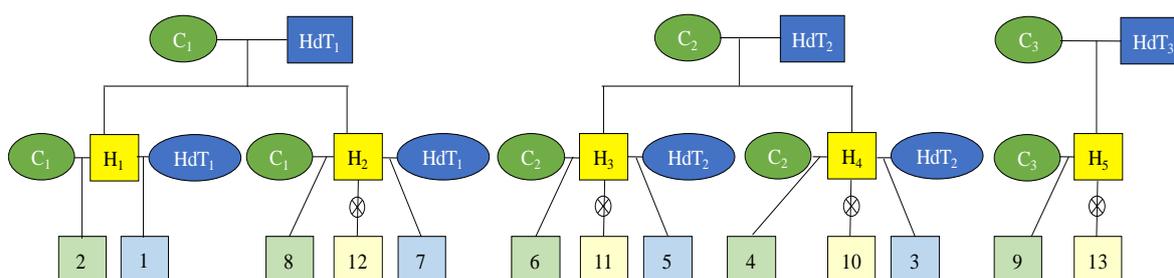
### **Índice de seleção**

O índice de seleção, rank-médio e ganho com a seleção estão apresentados na tabela 6. Este índice resulta da combinação das características preconizadas pelo melhorista, nas quais se deseja exercer a seleção simultânea. Nessas análises, foram utilizadas as características com efeito genético significativo pelo teste da razão de verossimilhança na ANADEV (Tabela 3). Conforme os índices de cada característica, foi calculado o rank médio pelo qual as progênies foram classificadas. Ao selecionar as cinco progênies (intensidade de seleção de 30%) por meio dos melhores ranks-médio, o ganho de seleção foi superior a 32%. Esse resultado demonstra a importância da metodologia estatística utilizada (BLUP) a qual corrige os dados para os efeitos ambientais e prediz de maneira precisa e não viciada os valores genotípicos, conduzindo à maximização do ganho genético com a seleção (Resende et al. 2001). Além disso, os resultados demonstraram que, assim como as progênies testemunhas, as cinco melhores progênies pelo índice de seleção (3, 5, 10, 11 e 12) foram

alocadas no grupo I do dendrograma (Figura 2). Deve-se ressaltar que, pelo índice de seleção, essas progênies foram superiores às progênies testemunhas, sendo, portanto, materiais promissores e que devem ser explorados nos programas de melhoramento genético dessa espécie. Em concordância com esses resultados, a superioridade de progênies obtidas de cruzamentos entre genitores Catuaí e HdT em relação a progênies Catuaí tem sido demonstrada (Bonomo et al. 2004).

### **Conclusão**

A população avaliada apresentou baixa variabilidade genética, demonstrando o alto grau de parentesco entre as progênies. Porém, ainda foi possível selecionar as progênies 3, 5, 10, 11 e 12 que proporcionaram ganhos com seleção acima de 32% e foram superiores às testemunhas utilizadas.



**Figura 1.** Heredograma das 13 progênies de *Coffea arabica* oriundas de cruzamentos entre genitores do grupo de cultivares Catuaí e Híbrido de Timor; C<sub>1</sub>, C<sub>2</sub> e C<sub>3</sub> = genótipos Catuaí amarelo IAC 30, IAC 86 e IAC 64, respectivamente; HdT<sub>1</sub>, HdT<sub>2</sub> e HdT<sub>3</sub> = genótipos Híbrido de Timor UFV 445-46, UFV 440-10 e UFV 530, respectivamente; H<sub>1</sub>, H<sub>2</sub>, H<sub>3</sub>, H<sub>4</sub> e H<sub>5</sub> = híbridos oriundos de cruzamentos entre genitores Catuaí amarelos e Híbrido de Timor; 1, 3, 5 e 7 = progênies de geração de retrocruzamento resistente; 2, 4, 6, 8 e 9 = progênies de geração de retrocruzamento suscetível; 10, 11, 12 e 13 = progênies em geração F<sub>2</sub>.

**Tabela 1.** Progenies de Coffea arabica avaliadas em Viçosa (MG)

<b>Progenies</b>	<b>Nº de plantas</b>	<b>Genitor 1</b>	<b>Genitor 2</b>
Catuaí Vermelho IAC 44	16	IAC 476-11	IAC 374-19
Oeiras MG 6851	17	CIFC 19/1	CIFC 832/1
RCr 1	15	H 419-1 c-17	UFV 445-46
RCs 2	15	H 419-1 c-17	UFV 2143-235
RCr 3	15	H 514-8 c-387	UFV 440-10
RCs 4	15	H 514-8 c-387	UFV 2154-344
RCr 5	15	H 514-7 c-364	UFV 440-10
RCs 6	15	H 514-7 c-364	UFV 2154-344
RCr 7	15	H 419-10 c-214	UFV 445-46
RCs 8	15	H 419-10 c-214	UFV 2143-235
RCs 9	15	UFV 2148-57	H 513-5 c-14
F <sub>2</sub> 10	15	H 514-8 c-387	-
F <sub>2</sub> 11	15	H 514-7 c-364	-
F <sub>2</sub> 12	15	H 419-10 c-214	-
F <sub>2</sub> 13	15	H 513-5 c-14	-

RCr = Retrocruzamento resistente; RCs = Retrocruzamento suscetível; F<sub>2</sub> = geração obtida pela autofecundação dos híbridos F<sub>1</sub>.

**Tabela 2.** Características fenotípicas avaliadas nos anos de 2014, 2015 e 2016 em Viçosa (MG)

<b>Característica</b>	<b>Forma de avaliação</b>
Produção	Litros de café recém colhido por planta
Comprimento de uma folha (cm)	Medidos em folha do 3º ou 4º par de um ramo plagiotrópicos do terço médio da planta
Largura de uma folha (cm)	
Comprimento de um ramo (cm)	Medidos no ramo plagiotrópico localizado no terço médio da planta
Número de nós reprodutivos	
Número de nós vegetativos	
Diâmetro do caule (cm)	Medido na região do coleto da planta (cerca de 5cm da superfície do solo)
Tamanho dos frutos maduros	Avaliado por notas de 1 a 3
Uniformidade de maturação dos frutos	Avaliada por nota de 1 a 4
Ciclo de maturação dos frutos	
Incidência de Ferrugem	Avaliada por notas de 1 a 5
Incidência de Cercosporiose	
Infestação de Bicho-Mineiro	
Vigor vegetativo	Avaliado por notas de 1 a 10

**Tabela 3.** Análise de deviance (ANADEV) e significâncias do teste de razão da verossimilhança (LRT) referentes à análise conjunta de 14 características morfológicas medidas nos anos de 2014, 2015 e 2016, no experimento de progênies de *Coffea arabica* em Viçosa, MG.

Variáveis	DEVIANCE					LRT (X <sup>2</sup> )			
	Modelo completo	Efeito entre progênies	Efeito permanente entre plantas	Efeito entre parcela	Efeito da interação progênies x ano	Efeito entre progênies	Efeito permanente entre plantas	Efeito entre parcela	Efeito da interação progênies x ano
Prod	1976,04	1984,45	2018,34	1978,88	1976,39	8,41**	42,3**	2,84 <sup>NS</sup>	0,35 <sup>NS</sup>
Fer	76,95	81,59	122,26	76,97	83,47	4,64*	45,31**	0,02 <sup>NS</sup>	6,52*
Cer	193,81	196,16	193,97	193,81	195,29	2,35 <sup>NS</sup>	0,16 <sup>NS</sup>	0 <sup>NS</sup>	1,48 <sup>NS</sup>
BM	-54,91	-53,69	-54,29	-49,81	-54,87	1,22 <sup>NS</sup>	0,62 <sup>NS</sup>	5,1*	0,04 <sup>NS</sup>
Vig	718,76	726,46	768,34	724,25	735,28	7,7**	49,58**	5,49*	16,52**
CMat	386,29	398,55	409,7	386,3	392,42	12,26**	23,41**	0,01 <sup>NS</sup>	6,13*
UMat	229,58	229,87	229,95	229,58	234,26	0,29 <sup>NS</sup>	0,37 <sup>NS</sup>	0 <sup>NS</sup>	4,68*
TFr	-291,1	-289,51	-264,78	-290,64	-288,76	1,59 <sup>NS</sup>	26,32**	0,46 <sup>NS</sup>	2,34 <sup>NS</sup>
CF	1424,08	1425,61	1430,88	1424,16	1429,13	1,53 <sup>NS</sup>	6,8**	0,08 <sup>NS</sup>	5,05*
LF	500,75	501,64	502,04	503,9	502,17	0,89 <sup>NS</sup>	1,29 <sup>NS</sup>	3,15 <sup>NS</sup>	1,42 <sup>NS</sup>
CRP	4100,98	4115,16	4136,99	4106,92	4102,7	14,18**	36,01**	5,94*	1,72 <sup>NS</sup>
NNR	2175,79	2179,09	2175,8	2176,03	2178,34	3,3 <sup>NS</sup>	0,01 <sup>NS</sup>	0,24 <sup>NS</sup>	2,55 <sup>NS</sup>
NNV	1999,75	2003,26	1999,79	2000,02	2000,29	3,51 <sup>NS</sup>	0,04 <sup>NS</sup>	0,27 <sup>NS</sup>	0,54 <sup>NS</sup>
DCa	532,94	533,98	542	538,49	550,02	1,04 <sup>NS</sup>	9,06**	5,55*	17,08**

\*\* e \* significativo a 1% e 5% de probabilidade, respectivamente, e <sup>NS</sup> não-significativo pelo teste qui-quadrado com 1 grau de liberdade. Prod = Produção de frutos de café; Fer = Incidência à ferrugem; Cer = Incidência à Cercosporiose; BM = Infestação de Bicho-Mineiro; Vig = Vigor vegetativo; CMat = Ciclo de maturação dos frutos; UMat = Uniformidade de maturação dos frutos; TFr = Tamanho dos frutos maduros; CF = Comprimento de uma folha; LF = Largura de uma folha; CRP = Comprimento de um ramo plagiotrópico; NNR = Número de nós reprodutivos; NNV = Número de nós vegetativos; DCa = Diâmetro do caule.

**Tabela 4.** Estimativa dos parâmetros genéticos referentes à análise conjunta de 14 características morfológicas medidas nos anos de 2014, 2015 e 2016, no experimento de progênies de *Coffea arabica* em Viçosa, MG.

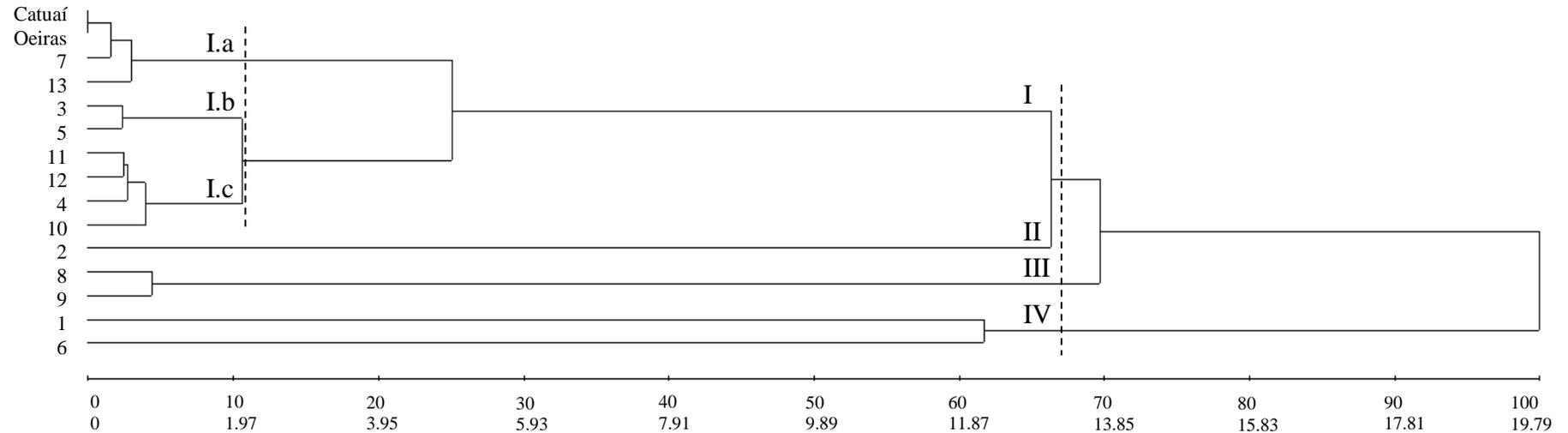
Parâmetros	Prod	Fer	Cer	BM	Vig	CMat	UMat	TFr	CF	LF	CRP	NNR	NNV	DCa
Vg	0,64	0,03	0,02	0,01	0,16	0,10	0,01	0,01	0,10	0,02	25,49	0,44	0,28	0,11
Vplan	1,97	0,12	0,01	0,01	0,31	0,13	0,01	0,05	0,31	0,03	40,68	0,07	0,06	0,20
Vparc	0,30	0,00	0,00	0,01	0,07	0,00	0,00	0,00	0,01	0,02	11,01	0,07	0,04	0,12
Vint	0,06	0,02	0,01	0,00	0,09	0,03	0,03	0,01	0,14	0,02	3,46	0,32	0,11	0,18
Ve	5,21	0,30	0,45	0,31	0,74	0,50	0,47	0,18	2,53	0,68	117,24	8,67	6,53	0,88
Vf	8,18	0,47	0,48	0,33	1,37	0,76	0,51	0,25	3,09	0,77	197,87	9,58	7,02	1,49
h <sup>2</sup> g	0,08	0,07	0,03	0,02	0,12	0,14	0,01	0,04	0,03	0,02	0,13	0,05	0,04	0,08
c <sup>2</sup> plan	0,24	0,25	0,01	0,03	0,23	0,17	0,01	0,20	0,10	0,04	0,21	0,01	0,01	0,14
c <sup>2</sup> parc	0,04	0,00	0,00	0,03	0,05	0,00	0,00	0,01	0,00	0,03	0,06	0,01	0,01	0,08
c <sup>2</sup> int	0,01	0,04	0,03	0,00	0,06	0,04	0,05	0,03	0,05	0,02	0,02	0,03	0,02	0,12
repet	0,36	0,32	0,04	0,07	0,39	0,31	0,02	0,25	0,14	0,09	0,39	0,06	0,05	0,29
Média	4,57	3,27	2,60	2,07	6,75	2,59	2,69	2,38	11,71	5,07	60,71	8,02	8,83	4,86
CVg %	17,47	5,66	4,77	3,49	5,93	12,46	2,82	4,05	2,67	2,45	8,32	8,30	5,96	6,92
CVe %	49,94	16,71	25,73	26,87	12,77	27,19	25,52	18,00	13,57	16,29	17,84	36,71	28,94	19,27
CVrel	0,35	0,34	0,19	0,13	0,46	0,46	0,11	0,22	0,20	0,15	0,47	0,23	0,21	0,36
rgg	0,67	0,66	0,60	0,48	0,78	0,85	0,40	0,56	0,55	0,48	0,72	0,62	0,63	0,54

Vg = variância entre progênies; Vplan = variância permanente entre plantas; Vparc = variância entre parcelas; Vint = variância da interação progênies X anos; Ve = variância ambiental; Vf = variância fenotípica; h<sup>2</sup>g = herdabilidade entre progênies; c<sup>2</sup>plan = coeficiente de determinação permanente entre plantas; c<sup>2</sup>parc = coeficiente de determinação entre parcelas; c<sup>2</sup>int = coeficiente de determinação da interação progênies X anos; repet = Repetibilidade; Média = Média geral; CVg % = Coeficiente de variação genotípico; CVe % = Coeficiente de variação residual; CVrel = Coeficiente de variação relativo; rgg total = Acurácia total. Prod = Produção de frutos de café; Fer = Incidência à ferrugem; Cer = Incidência à Cercosporiose; BM = Infestação de Bicho-Mineiro; Vig = Vigor vegetativo; CMat = Ciclo de maturação dos frutos; UMat = Uniformidade de maturação dos frutos; TFr = Tamanho dos frutos maduros; CF = Comprimento de uma folha; LF = Largura de uma folha; CRP = Comprimento de um ramo plagiotrópico; NNR = Número de nós reprodutivos; NNV = Número de nós vegetativos; DCa = Diâmetro do caule.

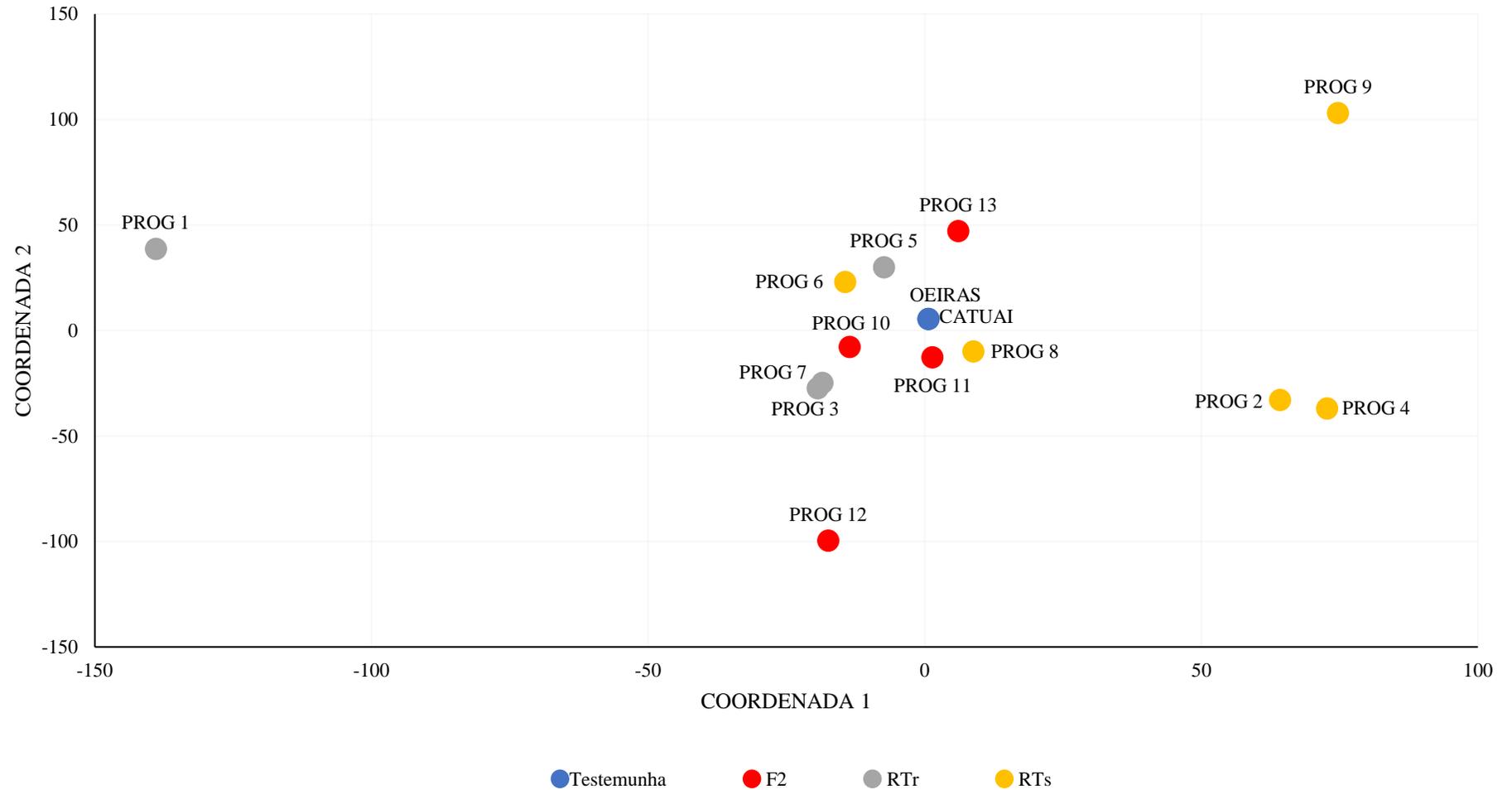
**Tabela 5.** Estimativa das correlações genéticas entre características referentes à análise conjunta de 14 características morfológicas medidas nos anos de 2014, 2015 e 2016, no experimento de progênies de Coffea arabica em Viçosa, MG.

VARIÁVEIS	Prod	Fer	Cer	BM	Vig	CMat	UMat	TFr	CF	LF	CRP	NNR	NNV	DCa
Prod	1,00	0,17	0,06	0,13	0,76	0,65	0,30	0,08	-0,14	-0,50	0,29	0,61	0,58	0,42
Fer		1,00	0,74	0,24	-0,15	-0,23	-0,16	0,31	-0,18	-0,17	-0,58	0,00	0,51	-0,32
Cer			1,00	0,44	-0,22	-0,24	-0,22	0,22	-0,42	-0,28	-0,73	-0,08	0,45	-0,41
BM				1,00	0,05	-0,08	-0,15	0,41	-0,22	-0,09	-0,24	0,19	0,12	0,43
Vig					1,00	0,80	0,54	-0,18	0,29	-0,05	0,37	0,57	0,46	0,47
CMat						1,00	0,75	-0,30	-0,08	-0,23	0,55	0,69	0,05	0,18
UMat							1,00	-0,58	0,09	-0,03	0,31	0,27	-0,08	-0,10
TFr								1,00	-0,07	0,05	-0,13	0,18	0,14	0,31
CF									1,00	0,83	0,01	-0,33	0,04	0,16
LF										1,00	-0,17	-0,46	-0,23	-0,08
CRP											1,00	0,50	-0,38	0,56
NNR												1,00	0,18	0,44
NNV													1,00	0,13
DCa														1,00

Prod = Produção de frutos de café; Fer = Incidência à ferrugem; Cer = Incidência à Cercosporiose; BM = Infestação de Bicho-Mineiro; Vig = Vigor vegetativo; CMat = Ciclo de maturação dos frutos; UMat = Uniformidade de maturação dos frutos; TFr = Tamanho dos frutos maduros; CF = Comprimento de uma folha; LF = Largura de uma folha; CRP = Comprimento de um ramo plagiotrópico; NNR = Número de nós reprodutivos; NNV = Número de nós vegetativos; DCa = Diâmetro do caule.



**Figura 2.** Dendrograma obtido pela técnica UPGMA (Unweighted Pair Group Method using Arithmetic averages) com base na matriz de distância genética euclidiana média a partir de 228 genótipos constituintes das progênes de *Coffea arabica* avaliada, conforme apresentado na tabela 1. CATUAÍ = Catuaí Vermelho IAC 44; OEIRAS = Oeiras MG 6851; 1, 3, 5 e 7 = progênes de geração de retrocruzamento resistente; 2, 4, 6, 8 e 9 = progênes de geração de retrocruzamento suscetível; 10, 11, 12 e 13 = progênes em geração F<sub>2</sub>.



**Figura 3.** Dispersão gráfica das coordenadas principais obtida com os dados da matriz de distância genética euclidiana média a partir de 228 genótipos constituintes das progênies de *Coffea arabica* avaliada, conforme apresentado na tabela 1. CATUAÍ = Catuaí Vermelho IAC 44; OEIRAS = Oeiras MG 6851.

**Tabela 6.** Índice de seleção, rank, rank-médio e ganho de seleção em porcentagem das progênies de cafeeiros obtidos a partir das características com efeito entre progênies significativo pela análise ANADEV, conforme a Tabela 3.

Ordem	Progênie	Prod	Fer	Vig	CMat	CRP	Rank-Médio	Ganho %
1	PROG 10	2	4	2	9	3	4,00	100,00
2	PROG 3	7	5	4	14	1	6,20	56,86
3	PROG 11	8	3	7	10	5	6,60	42,86
4	PROG 12	12	1	3	13	4	6,60	36,75
5	PROG 5	4	6	10	12	2	6,80	32,45
6	PROG 2	1	13	1	11	11	7,40	27,66
7	PROG 1	15	2	15	1	6	7,80	23,35
8	PROG 7	11	8	5	6	9	7,80	20,30
9	T2-CATUAI	9	9	8	7	7	8,00	17,65
10	PROG 13	5	11	12	3	12	8,60	14,61
11	PROG 4	6	7	6	15	10	8,80	11,96
12	T1-OEIRAS	10	10	9	8	8	9,00	9,59
13	PROG 9	3	15	13	4	13	9,60	7,00
14	PROG 6	14	12	14	2	15	11,40	3,13
15	PROG 8	13	14	11	5	14	11,40	0,00

Prod = Produção de frutos de café; Fer = Incidência à ferrugem; Cer = Incidência à Cercosporiose; BM = Infestação de Bicho-Mineiro; Vig = Vigor vegetativo; CMat = Ciclo de maturação dos frutos; UMat = Uniformidade de maturação dos frutos; TFr = Tamanho dos frutos maduros; CF = Comprimento de uma folha; LF = Largura de uma folha; CRP = Comprimento de um ramo plagiotrópico; NNR = Número de nós reprodutivos; NNV = Número de nós vegetativos; DCa = Diâmetro do caule.

## Referências Bibliográficas

- Alkimim ER, Caixeta ET, Sousa TV, et al (2017) Marker-assisted selection provides arabica coffee with genes from other *Coffea* species targeting on multiple resistance to rust and coffee berry disease. *Molecular Breeding* 37:6. doi: 10.1007/s11032-016-0609-1
- Andrew RL, Wallis IR, Harwood CE, Foley WJ (2010) Genetic and environmental contributions to variation and population divergence in a broad-spectrum foliar defence of *Eucalyptus tricarpa*. *Annals of Botany* 105:707–717. doi: 10.1093/aob/mcq034
- Barbosa MHP, Resende MDV, Dias LA dos S, et al (2012) Genetic improvement of sugar cane for bioenergy: the brazilian experience in network research with RIDESA. *Crop Breeding and Applied Biotechnology* 12:87–98. doi: 10.1590/S1984-70332012000500010
- Bertrand B, Anthony F, Lashermes P (2001) Breeding for resistance to *Meloidogyne exigua* in *Coffea arabica* by introgression of resistance genes of *Coffea canephora*. *Plant Pathology* 50:637–643. doi: 10.1046/j.1365-3059.2001.00597.x
- Bonomo P, Cruz CD, Viana JMS, et al (2004) Avaliação de progênies obtidas de cruzamentos de descendentes do Híbrido de Timor com as cultivares Catuaí Vermelho e Catuaí Amarelo. *Bragantia* 63:207–219. doi: 10.1590/S0006-87052004000200006
- Carvalho A, Medina Filho HP, Fazuoli LC, Costa WM da (1984) Genética de *Coffea*: XXVI. Hereditariedade do porte reduzido do cultivar Caturra. *Bragantia* 43:443–458. doi: 10.1590/S0006-87051984000200015
- Corrêa TR, Motoike SY, Coser SM, et al (2015) Estimation of genetic parameters for in vitro oil palm characteristics (*Elaeis guineensis* Jacq.) and selection of genotypes for cloning capacity and oil yield. *Industrial Crops and Products* 77:1033–1038. doi: 10.1016/j.indcrop.2015.09.066
- Cruz CD (2013) GENES - a software package for analysis in experimental statistics and quantitative genetics. *Acta Scientiarum Agronomy* 35:271–276. doi: 10.4025/actasciagron.v35i3.21251
- Cruz CD, Regazzi AJ, Carneiro PCS (2004) Modelos biométricos aplicados ao melhoramento genético, Editora UF. viçosa
- Cubry P, Musoli P, Legnate H, et al (2008) Diversity in coffee assessed with SSR markers: Structure of the genus *Coffea* and perspectives for breeding. *Genome* 51:50–63. doi: 10.1139/G07-096

- Diola V, de Brito GG, Caixeta ET, et al (2011) High-density genetic mapping for coffee leaf rust resistance. *Tree Genetics & Genomes* 7:1199–1208. doi: 10.1007/s11295-011-0406-2
- Falconer DS (1987) *Introdução à genética quantitativa*, UFV. viçosa
- Fernandes ALT, Partelli FL, Bonomo R, Golynski A (2012) A moderna cafeicultura dos cerrados brasileiros. *Pesquisa Agropecuária Tropical* 42:231–240. doi: 10.1590/S1983-40632012000200015
- Ferreira RT, Viana AP, Barroso DG, et al (2012) Toona ciliata genotype selection with the use of individual BLUP with repeated measures. *Scientia Agricola* 69:210–216. doi: 10.1590/S0103-90162012000300006
- Gichimu B, Gicheru E, Mamati G, Nyende A (2013) Variation and association of cup quality attributes and resistance to Coffee Berry Disease in *Coffea arabica* L. composite cultivar, Ruiru 11. *African Journal of Horticultural Science* 7:22–35.
- Gichimu BM, Gichuru EK, Mamati GE, Nyende AB (2014) Occurrence of Ck-1 gene conferring resistance to Coffee Berry Disease in *Coffea arabica* cv. Ruiru 11 and its parental genotypes. *Journal of Agricultural and Crop Research* 2:51–61.
- Henderson CR (1975) Best linear unbiased estimation and prediction under a selection model. *Biometrics* 31:423–449.
- Hoffmann AA, Shirriffs J, Scott M (2005) Relative importance of plastic vs genetic factors in adaptive differentiation: geographical variation for stress resistance in *Drosophila melanogaster* from eastern Australia. *Functional Ecology* 19:222–227. doi: 10.1111/j.1365-2435.2005.00959.x
- Inácio P, Lewinsohn T, Carmo RL do, Hogan DJ (2002) Ordenação multivariada na ecologia e seu uso em ciências ambientais. *Ambiente & Sociedade* 69–83. doi: 10.1590/S1414-753X2002000100005
- Lashermes P, Combes M-C, Robert J, et al (1999) Molecular characterisation and origin of the *Coffea arabica* L. genome. *Molecular and General Genetics MGG* 261:259–266. doi: 10.1007/s004380050965
- Lashermes P, Combes MC, Ansaldi C, et al (2011) Analysis of alien introgression in coffee tree (*Coffea arabica* L.). *Molecular Breeding* 27:223–232. doi: 10.1007/s11032-010-9424-2
- Monteiro JEBA, Sentelhas PC, Chiavegato EJ, et al (2005) Estimação da área foliar do algodoeiro por meio de dimensões e massa das folhas. *Bragantia* 64:15–24. doi: 10.1590/S0006-87052005000100002

- Mrode R (ed) (2014) Linear models for the prediction of animal breeding values. CABI, Wallingford
- Mulamba NN, Mock JJ (1978) Improvement of yield potential of the Eto Blanco maize (*Zea mays* L.) population by breeding for plant traits. *Journal of Genetics and Cytology* 7:40–51.
- Patterson HD, Thompson R (1971) Recovery of inter-block information when block sizes are unequal. *Biometrika* 58:545–554. doi: 10.1093/biomet/58.3.545
- Pereira AA, Carvalho GR, Moura WM, et al (2010a) Cultivares: Origem e suas Características. In: Reis PR, Cunha RL (eds) *Café arábica do plantio à colheita*, EPAMIG. Lavras, pp 167–221
- Pereira JCC (2008) *Melhoramento genético aplicado à produção animal*. FEPMVZ Editora, Belo Horizonte
- Pereira MC, Chalfoun SM, Carvalho GR de, Savian TV (2010b) Multivariate analysis of sensory characteristics of coffee grains (*Coffea arabica* L.) in the region of upper Paranaíba. *Acta Scientiarum Agronomy*. doi: 10.4025/actasciagron.v32i4.4283
- Pereira TB, Carvalho JPF, Botelho CE, et al (2013) Eficiência da seleção de progênes de café F4 pela metodologia de modelos mistos (REML/BLUP). *Bragantia* 72:230–236. doi: 10.1590/brag.2013.031
- Piepho HP, Möhring J, Melchinger AE, Büchse A (2008) BLUP for phenotypic selection in plant breeding and variety testing. *Euphytica* 161:209–228. doi: 10.1007/s10681-007-9449-8
- Resende MDV (2002) *Genética biométrica e estatística no melhoramento de plantas perenes*, EMBRAPA In. Brasília
- Resende MDV, Duarte JB (2007) Precisão e controle de qualidade em experimentos de avaliação de cultivares. *Pesquisa Agropecuária Tropical* 37:182–194.
- Resende MDV, Thompson R (2004) Factor analytic multiplicative mixed models in the analysis of multiple experiments. *Revista de Matemática e Estatística* 22:1–22.
- Resende MAV de, Freitas JA de, Lanza MA, et al (2014a) Divergência genética e índice de seleção via BLUP em acessos de algodoeiro para características tecnológicas da fibra. *Pesquisa Agropecuária Tropical* 44:334–640. doi: 10.1590/S1983-40632014000300006
- Resende MDV de (2007) *Matemática e estatística na análise de experimentos e no melhoramento genético*. Colombo: Embrapa Florestas
- Resende MDV de (2016) *Software Selegen-REML/BLUP: a useful tool for plant breeding*.

- Crop Breeding and Applied Biotechnology 16:330–339. doi: 10.1590/1984-70332016v16n4a49
- Resende MDV de (2015) *Genética Quantitativa e de Populações*, 1st edn. Viçosa
- Resende MDV De, Furlani-Júnior E, Moraes MLT De, Fazuoli LC (2001) Estimativas de parâmetros genéticos e predição de valores genotípicos no melhoramento do cafeeiro pelo procedimento REML/BLUP. *Bragantia* 60:185–193. doi: 10.1590/S0006-87052001000300005
- Resende MDV de, Silva FF e, Azevedo CF (2014b) *Estatística Matemática, Biométrica e Computacional: Modelos Mistos, Multivariados, Categorias e Generalizados (REML/BLUP), Inferência Bayesiana, Regressão Aleatória, Seleção Genômica, QTI-GWAS, Estatística Espacial e Temporal, Competição, Sobrevivência*, 1st edn. Suprema, Viçosa
- Romero G, Vásquez LM, Lashermes P, Herrera JC (2014) Identification of a major QTL for adult plant resistance to coffee leaf rust (*Hemileia vastatrix*) in the natural Timor hybrid (*Coffea arabica* x *C. canephora*). *Plant Breeding* 133:121–129. doi: 10.1111/pbr.12127
- Setotaw TA, Caixeta ET, Pereira AA, et al (2013) Coefficient of Parentage in *Coffea arabica* L. Cultivars Grown in Brazil. *Crop Science* 53:1237–1247. doi: 10.2135/cropsci2012.09.0541
- Silvarolla MB, Guerreiro Filho O, Lima MMA de, Fazuoli LC (1997) Avaliação de progênes derivadas do híbrido de timor com resistência ao agente da ferrugem. *Bragantia* 56:47–58. doi: 10.1590/S0006-87051997000100005
- Sobreira FM, Oliveira ACB de, Pereira AA, et al (2015) Sensory quality of arabica coffee (*Coffea arabica*) genealogic groups using the sensogram and content analysis. *Australian Journal of Crop Science* 9:486–493.
- Sousa TVTV, Caixeta ETET, Alkimim ERER, et al (2017) Molecular markers useful to discriminate *Coffea arabica* cultivars with high genetic similarity. *Euphytica* 213:75. doi: 10.1007/s10681-017-1865-9
- Spinelli VM, Dias LAS, Rocha RB, Resende MD V. (2015) Estimates of genetic parameters with selection within and between half-sib families of *Jatropha curcas* L. *Industrial Crops and Products* 69:355–361. doi: 10.1016/j.indcrop.2015.02.024
- Teixeira AL, Gonçalves FMA, Rezende JC de, et al (2012) Seleção precoce para produção de grãos em café arábica pela avaliação de caracteres morfológicos. *Pesquisa Agropecuária Brasileira* 47:1110–1117. doi: 10.1590/S0100-204X2012000800011
- Vencovsky R (1987) Herança quantitativa. In: PATERNIANI E, VIEGAS GP (eds)

Melhoramento e produção de milho, 2nd edn. Fundação Cargill, Campinas, pp 137–214

Viana JMS, Faria VR, Silva FF e, Resende MDV de (2011) Best Linear Unbiased Prediction and Family Selection in Crop Species. *Crop Science* 51:2371. doi: 10.2135/cropsci2011.03.0153

### **CAPÍTULO 3**

#### **Seleção genômica ampla em população de Coffea arabica**

VIÇOSA  
MINAS GERAIS - BRASIL  
2017

## RESUMO

O café arábica (*Coffea arabica*) é a principal espécie comercial do gênero *Coffea*. Embora a seleção genômica Ampla (GWS) tenha se mostrado altamente eficiente para diversas espécies vegetais anuais e perenes, em café arábica nenhum estudo de GWS foi publicado. Assim, objetivou-se avaliar a acurácia seletiva e a eficiência da GWS utilizada para prever o valor genético genômico (GEBV) dos indivíduos que compõem a população de cafeeiro arábica obtida por cruzamentos entre genitores do grupo de cultivares Catuaí e Híbrido de Timor. Um total de 195 indivíduos de café arábica, pertencentes a 13 famílias em geração F<sub>2</sub>, retrocruzamentos suscetíveis e retrocruzamentos resistentes à ferrugem do cafeeiro foram fenotipados para 18 características agrônomicas e genotipados com 21.211 marcadores moleculares SNP. Os dados fenotípicos foram analisados por meio dos modelos mistos (REML/BLUP), implementados no software Selegen-REML/BLUP. As análises de GWS foram realizadas utilizando o método G-BLUP via procedimento RKHS (Reproducing Kernel Hilbert Spaces) com algoritmo Bayesiano, por meio do pacote BGLR, implementado no software R. As herdabilidades fenotípicas e as acurácias seletivas, com base nos dados fenotípicos, foram estimadas, as quais foram de magnitude moderada a alta, para a maioria das características avaliadas. Os resultados das análises de GWS foram promissores, sendo esta mais eficiente que a seleção utilizando, exclusivamente, os dados fenotípicos e indicaram a possibilidade de se maximizar os ganhos com seleção por unidade de tempo. O efeito da densidade de marcadores nas análises de GWS foi avaliado. Foi possível concluir que a seleção genômica ampla mostrou-se potencial para o melhoramento de *C. arabica* por possibilitar a obtenção acurada dos GEBV e pela maior eficiência em relação a seleção tradicional possibilitando a redução no tempo necessário para completar o ciclo de seleção.

**Palavras chave:** Café, GWS, capacidade preditiva, acurácia seletiva.

## **Introdução**

O melhoramento genético vegetal iniciou de forma empírica com a seleção fenotípica de indivíduos que se destacavam positivamente nas populações segregantes. Nos anos 80, marcadores moleculares foram desenvolvidos e passaram a ser utilizados como ferramenta auxiliar às informações fenotípicas (Soller and Beckmann 1983). Com a evolução da biologia molecular, nos anos 90, a Seleção Assistida por Marcadores Moleculares (SAM) foi idealizada (Lande and Thompson 1990) e possibilitou a seleção de indivíduos portadores de alelos específicos. No entanto, a SAM mostrou-se pouco eficiente em características poligênicas e ou de baixa herdabilidade (Bernardo 2008). Essa limitação é devido, principalmente, ao fato dos marcadores moleculares em associações significativas com os QTL (Quantitative Trait Loci) serem incapazes de capturarem genes de efeito menor (Heffner et al. 2009; Hayes et al. 2009; Xu et al. 2012).

Por vislumbrar o potencial e a importância da mesma, a seleção genômica ampla (GWS) foi proposta (Meuwissen et al. 2001) e, atualmente, é amplamente utilizada em animais e vegetais. A rápida adoção dessa técnica seletiva se deve, entre outros fatores, ao fato de combinar número expressivo de marcadores cobrindo todo genoma da espécie. Além disso, essa técnica está aliada a metodologias estatísticas robustas e acuradas capazes de estimar valores genéticos dos indivíduos (Longin et al. 2015) aumentando os ganhos com seleção por unidade de tempo (Heffner et al. 2010). Vários estudos demonstram sua alta acurácia seletiva (Bernardo and Yu 2007; Wong and Bernardo 2008; Heffner et al. 2009; Crossa et al. 2010; Garcia et al. 2011; Grattapaglia and Resende 2011; Davey et al. 2011; Iwata et al. 2011; Resende et al. 2012a; Resende et al. 2012b; De Los Campos et al. 2013; Gianola 2013). Além disso, a GWS tem se mostrado eficiente para caracteres poligênicos, com baixa herdabilidade, de difícil mensuração e avaliação dispendiosa (Heslot et al. 2015; Poland 2015).

Com o aperfeiçoamento das plataformas de sequenciamento (Next Generation Sequencing - NGS), a GWS tornou-se realidade para diversas espécies de importância econômica. Com o uso das plataformas de NGS, novos marcadores foram economicamente viabilizados (Patel et al. 2015), dentre os quais se destacam os marcadores SNP (Single Nucleotide Polymorphisms). SNP é a variação genética mais abundantes no genoma (Kwok and Gu 1999; Ganal et al. 2009) e, dessa forma, possibilita a identificação de milhares desses marcadores distribuídos ao longo do genoma das espécies, os quais podem ser utilizados nas análises preditivas de seleção genômica.

O uso de marcadores moleculares SNP em estudos de GWS tem se mostrado vantajoso para muitas espécies, no entanto, cuidados especiais devem ser considerados ao se utilizar esses marcadores na espécie *Coffea arabica*. Essa espécie é originada do cruzamento natural, a partir de gametas não reduzidos entre as espécies diploides *Coffea canephora* e *Coffea eugenioides* (Lashermes et al. 1999), as quais apresentam genomas com regiões de alta similaridade (Cenci et al. 2012). Embora *C. arabica* seja alotetraploide verdadeiro (Clarindo and Carvalho 2008), exibe comportamento meiótico como diploide com formação dos bivalente (Lashermes et al. 2016). O genoma de *C. arabica* é composto por dois subgenomas, os quais podem apresentar regiões duplicadas. Assim, se o polimorfismo detectado pelo SNP ocorrer entre essas regiões dos subgenomas, esse marcador não explicará a variação fenotípica observada entre os indivíduos sendo, portanto, não informativo e denominado falso SNP (Vidal et al. 2010). Dessa forma, este deve ser eliminado do conjunto de dados. Além disso, o número ótimo de marcadores moleculares utilizados na predição dos valores genéticos dos indivíduos deve ser almejado. Marcadores em excesso associados a reduzido número de observações (genótipos), podem acarretar em problemas de multicolinearidade e de dimensionamento. Assim, é importante que as análises de qualidade

dos SNP corrijam ao máximo essas limitações maximizando a estimativa das acurácias preditivas.

Em plantas perenes, como *C. arabica*, a aplicabilidade da GWS assume importância acentuada (Resende et al. 2012c). Nessa espécie, a seleção dos indivíduos superiores tem sido realizada, basicamente, por meio de análises que utilizam, sobretudo, dados fenotípicos de produção e resistência às doenças. A metodologia de modelos mistos, REML/BLUP - Residual or Restricted Maximum Likelihood/Best Linear Unbiased Prediction (Patterson and Thompson 1971; Henderson 1975) tem sido aplicada com sucesso em vegetais (Resende and Thompson 2004; Viana et al. 2011; Ferreira et al. 2012; Barbosa et al. 2012; Pereira et al. 2013; Spinelli et al. 2015; Corrêa et al. 2015) por permitir, a partir de informações fenotípicas, a predição acurada e não viesada dos valores genéticos de indivíduos. Progressos genéticos também tem sido relatados pela utilização da seleção assistida por marcadores moleculares (Alkimim et al. 2017; Favoretto et al. 2017). Contudo, embora apresente grande importância econômica, nenhum trabalho de GWS foi publicado para essa espécie. Devido à complexidade e o número de genes que governam a maioria das características agronômicas dessa espécie, estudos de GWS são importantes por permitir o acesso às informações do genoma e por possibilitar a captura dos efeitos de todos os loci que explicam a variação genética (Heffner et al. 2009).

Diante do exposto, objetivou-se avaliar a acurácia seletiva e a eficiência da GWS utilizada para prever o valor genético genômico de população de cafeeiro arábica obtida por cruzamentos entre genitores do grupo de cultivares Catuaí e Híbrido de Timor.

## **Material e métodos**

### **Condução experimental**

Na área experimental, foram realizadas a calagem do solo e a adubação de plantio, conforme exigências da cultura. Os genótipos foram plantados no dia 11 de fevereiro de

2011. As plantas foram dispostas em espaçamento de 3,0 m entre fileiras e 0,7 m entre plantas. Nenhum método de controle fitossanitário para controle de ferrugem, antracnose e bicho mineiro foram aplicados. As avaliações do experimento foram realizadas em campo experimental do Departamento de Fitopatologia da Universidade Federal de Viçosa, em Viçosa, MG (20°44'25" S; 42°50'52" W), nos anos de 2014, 2015 e 2016.

### **Material genético**

A partir de cruzamentos entre três genitores do grupo de cultivares Catuaí e três genitores Híbrido de Timor (HdT), contrastantes em relação à ferrugem do cafeeiro, 13 progênies foram obtidas no programa de melhoramento genético de *C. arabica* da Epamig/UFV/Embrapa (Figura 1). Estas progênies são gerações de retrocruzamentos resistentes (RCr), retrocruzamentos suscetíveis (RCs) e F<sub>2</sub> (Figura 1 e Tabela 1). Em cada progênie foram selecionados 15 genótipos, totalizando 195 indivíduos.

### **Avaliações Fenotípicas**

As avaliações fenotípicas de 18 características agrônômicas foram realizadas (Tabela 2), onze contínuas e sete categóricas, nos 195 genótipos de *C. arabica* relacionados na Tabela 1.

As características contínuas foram avaliadas a partir de mensurações, conforme descrito na tabela 2. Já as características categóricas, conforme detalhado a seguir, foram avaliadas por notas. O tamanho dos frutos maduros foi avaliado com notas de 1, 2 e 3 para frutos pequenos, médios e graúdos, respectivamente. A uniformidade de maturação foi avaliada por notas de 1 a 4, sendo as notas 1, 2, 3 e 4 para uniforme, semiuniforme, semidesuniforme e desuniforme, respectivamente. O ciclo de maturação dos frutos foi avaliado com notas de 1 a 5, sendo notas 1, 2, 3, 4 e 5 para precoce, semiprecoce, médio, semitardio e tardio, respectivamente. As incidências à ferrugem do cafeeiro, à cercosporiose

e ao bicho mineiro foram avaliadas com notas de 1 a 5, onde as notas 1 e 5 foram atribuídas para genótipos sem nenhum sintoma das enfermidades e para genótipos altamente suscetíveis às mesmas, respectivamente. A característica vigor vegetativo foi avaliada por notas de 1 a 10, onde a nota 1 foi atribuída a plantas totalmente depauperada e a nota 10 para plantas consideradas de vigor vegetativo máximo.

### **Análises fenotípicas**

Os dados fenotípicos foram corrigidos e utilizados para estimar as acurácias seletivas ( $r_{yy}$ ) e herdabilidades fenotípicas ( $h^2_{fen}$ ) das 18 características agronômicas. Os dados foram corrigidos para efeitos de anos, parcelas e interação anos x parcelas. As análises foram realizadas considerando os modelos lineares mistos (procedimento REML/BLUP), por meio do software Selegen-REML/BLUP (Resende 2016). A estimação dos parâmetros genéticos foi realizada pela análise individual dos 18 caracteres utilizando o seguinte modelo estatístico:

$$y = Xu + Zg + Wp + Vr + Tb + Ri + e$$

Em que,

$y$  é o vetor de dados;

$u$  é o vetor referente à média geral em cada ano de avaliação;

$g$  é o vetor de efeitos de progênies (efeito aleatório);

$p$  é variância permanente entre plantas (efeito aleatório);

$r$  é variância entre tipos de populações (efeito aleatório);

$b$  é variância entre parcela (efeito aleatório);

$i$  é variância da interação progênies X anos (efeito aleatório);

$e$  é o vetor de resíduos (efeito aleatório).

As letras maiúsculas representam as matrizes de incidência para os referidos efeitos.

### **Extração do DNA genômico**

Folhas jovens e completamente expandidas dos 195 genótipos foram coletadas e o DNA genômico extraído utilizando a metodologia descrita por (Diniz et al. 2005). A concentração do DNA foi verificada em NanoDrop 2000 e sua qualidade avaliada em gel de agarose 1%.

A concentração de DNA das amostras foi padronizada e as mesmas enviadas para a empresa Rapid Genomics, Flórida/EUA, para construção das sondas, sequenciamento e identificação dos marcadores moleculares SNP.

### **Análise de qualidade de marcadores moleculares**

A partir da construção de 40.000 sondas polimórficas (Resende et al. 2016), os 195 cafeeiros foram genotipados com 21.211 marcadores SNPs. O conjunto de marcadores moleculares SNP foi submetido a análise de qualidade sendo removido os SNP que não atenderam os critérios mínimos estabelecidos nessa análise. As análises de qualidade foram conduzidas no software Rbio (Bhering 2017). Os parâmetros de qualidade utilizados foram CR (Call Rate) e MAF (Minor Allele Frequency) igual ou superior a 90% e a 5%, respectivamente. O nível crítico para o parâmetro MAF foi obtido pela equação  $MAF = \frac{1}{\sqrt{2N}}$ , onde N refere-se ao número de indivíduos avaliados. Além disso, para evitar a ocorrência de SNP falsos (Vidal et al. 2010), os marcadores moleculares que não apresentaram variância entre os indivíduos que compõem a população em estudo foram eliminados.

### Validação Cruzada

A validação cruzada é um método empregado para avaliar a capacidade de generalização de um modelo preditivo, a partir de um conjunto de dados. Na aplicação do método, o conjunto de dados é particionado em subconjuntos mutuamente exclusivos. A população, composta por 195 cafeeiros, foi dividida em 13 partes (folds), sendo 180 indivíduos utilizados para o treinamento ou estimação dos modelos preditivos e 15 genótipos para a validação dos mesmos. O processo foi repetido por 13 vezes, de forma que cada parte foi usada uma vez como conjunto de validação. Ao final, foi estimada a capacidade preditiva ( $r_{gy}$ ) obtida pelo resultado da média das correlações entre os valores genéticos genômicos estimados e os valores fenotípicos observados.

### Método de seleção genômica

As análises de seleção genômica foram realizadas utilizando o método G-BLUP via procedimento RKHS (Reproducing Kernel Hilbert Spaces) com algoritmo Bayesiano (Gianola 2006), por meio do pacote BGLR (Perez and de los Campos 2014), implementado no software R (R Core Team 2017).

O modelo linear misto geral (Resende 2007; Resende 2008) foi ajustado para estimar os efeitos dos marcadores:  $y = Xb + Wm + e$ , em que  $y$  é o vetor de observações fenotípicas,  $b$  é o vetor de efeitos fixos,  $m$  é o vetor dos efeitos aleatórios de marcadores e  $e$  refere-se ao vetor de resíduos aleatórios. As letras maiúsculas representam as matrizes de incidência para os referidos efeitos. A matriz de incidência  $X$  contém os valores 0, 1 e 2 para o número de alelos do marcador (ou do suposto QTL) em um indivíduo diploide. As equações de modelo misto genômicas para a predição de  $m$  via o método GBLUP equivalem a:

$$\begin{bmatrix} X'X & X'W \\ W'X & W'W + I \frac{\sigma_e^2}{(\sigma_g^2/n_Q)} \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{m} \end{bmatrix} = \begin{bmatrix} X'y \\ W'y \end{bmatrix}$$

O valor genético genômico global do indivíduo  $j$  é dado por  $VGG = \sum_i w_{ij} \hat{m}_i$ , em que  $W_i$  equivale a 0, 1 ou 2 para os genótipos mm, Mm e MM, respectivamente, para o marcador bialélico e codominante  $i$  (SNP). O componente  $W_{ij}$  refere-se ao elemento  $i$  da linha  $j$  da matriz  $W$ , referente ao indivíduo  $j$ .

### **Capacidade preditiva, acurácia da GWS e $n_{QTL}$**

A capacidade preditiva ( $r_{gy}$ ) é obtida por meio da correlação entre os valores genômicos preditos e os valores fenotípicos corrigidos, sendo equivalente à capacidade preditiva da GWS em estimar os fenótipos (Resende et al. 2014b).

A acurácia foi obtida pelo estimador:  $r_{gg} = \frac{r_{gy}}{\sqrt{h^2}}$ , em que  $r_{gy}$  equivale a capacidade preditiva da GWS e  $h^2$  a herdabilidade individual (Legarra et al. 2008).

A estimativa do número de QTL que controlam cada característica foi estimada a partir da expressão:  $n_{QTL} = \frac{(1-r_{gg}^2)Nh^2}{r_{gg}^2}$ , em que  $r_{gg}$  equivale a acurácia da GWS;  $N$  ao número de indivíduos da população e  $h^2$  a herdabilidade individual (Resende et al. 2008).

A estimativa do número de indivíduos que devem ser avaliados para se obter uma acurácia desejada foi obtida pela expressão:  $Ni = \frac{r_{gg}^2 n_{QTL}}{(1-r_{gg}^2)h^2}$ , em que  $r_{gg}$  equivale a acurácia da GWS;  $n_{QTL}$  ao número de QTLs que controlam cada característica e  $h^2$  a herdabilidade individual (Resende et al. 2014b).

### **Densidade de marcadores**

O efeito do número de marcadores na acurácia seletiva foi avaliado. A acurácia preditiva, com diferentes densidades de marcadores SNP, foi estimada pelo método G-BLUP via procedimento RKHS (Reproducing Kernel Hilbert Spaces) com algoritmo Bayesiano (Gianola 2006), por meio do pacote BGLR (Perez and de los Campos 2014), implementado

no software R (R Core Team 2017). Essas análises foram realizadas com conjunto de marcadores composto por 1.000, 4.000, 8.000, 12.000, 16.000, 20.000 e 20.477 SNP selecionados, aleatoriamente, de forma a amostrar o conjunto de dados original. Nessas análises, utilizou-se a validação cruzada utilizando 13 folds.

### **Eficiência da GWS**

A eficiência seletiva da GWS em comparação com a seleção baseada apenas em fenótipos, foi calculada usando a expressão:  $Ef = \frac{r_{gg}L_f}{r_{yy}L_{GWS}}$ , em que  $r_{gg}$  é acurácia seletiva da GWS;  $L_f$  é o tempo médio requerido para o ciclo de seleção com base em fenótipos;  $r_{yy}$  é a acurácia da seleção fenotípica;  $L_{GWS}$  é o tempo médio requerido para o ciclo de seleção com base na GWS (Resende et al. 2012d). As análises de eficiências foram estimadas considerando a necessidade de seis anos para obtenção das acurácias fenotípicas, conforme tempo de obtenção dos dados fenotípicos. Para a obtenção das acurácias da GWS foram considerados os tempos de um a seis anos, conforme possibilidade da técnica.

### **Resultados e discussão**

Os 195 cafeeiros, pertencentes às 13 famílias obtidas de cruzamentos entre genitores do grupo de cultivares Catuaí e Híbrido de Timor (HdT), tiveram 18 características de importância agrônômicas fenotipadas. A partir das informações fenotípicas, por meio da metodologia de modelos mistos (REML/BLUP), as herdabilidades fenotípicas ( $h^2_{fen}$ ) e as acurácias seletivas baseada nos dados fenotípicos ( $r_{yy}$ ) foram estimadas (Tabela 3). A característica diâmetro do caule (DCa) apresentou a menor estimativa de  $h^2_{fen}$  (0,01), ao passo que as características altura de plantas (APL) e diâmetro da copa (DCo) apresentaram os maiores valores estimados para esse parâmetro (0,90). As estimativas de  $h^2_{fen}$  para a maioria das características foram consideradas de magnitude média a alta (Resende and

Duarte 2007). A maioria das características avaliadas apresentaram  $r_{yy}$  consideradas de alta magnitude (Resende and Duarte 2007) (Tabela 3). Apenas a característica DCa apresentou estimativas de  $r_{yy}$  de baixa magnitude (Resende and Duarte 2007).

Os cafeeiros, além de fenotipados, foram genotipados a partir de 21.211 marcadores SNP. Após as análises de qualidades, 20.477 SNP foram selecionados, sendo observado redução de 3,46% do total de SNP que compunham o conjunto inicial de marcadores (Figura 2). A maior redução, em percentagem, do número de marcadores foi observada no cromossomo 4, a qual correspondeu a 14,21%. Os marcadores foram identificados em todos os cromossomos da espécie, amplamente distribuídos. O número de SNP por cromossomo variou de 49 (UNIGENE) a 2.804 (cromossomo 2), sendo observado média de 1.575 SNP por cromossomo. Em estudos com dados simulados foi demonstrado que a utilização de alta quantidade de marcadores possibilitou reduzir a limitação imposta pelo pequeno tamanho das populações de treinamento (Resende 2008). O aumento da densidade dos marcadores garante a conservação das associações marcador-QTL e possibilita obtenção de altas acurácias seletivas (Desta and Ortiz 2014). A densidade do marcador é determinada, principalmente, pela extensão do desequilíbrio de ligação (LD) e tamanho da amostra. Assim, se o número marcadores utilizados forem reduzidos, o tamanho populacional deve ser aumentado (Grattapaglia and Resende 2011).

Os valores estimados de herdabilidade genômica  $h^2_a$  variaram de 0,16 para a característica diâmetro do caule (DCa) a 0,46 para as características número de nós vegetativos (NNV) e altura de plantas (API) (Tabela 3). As estimativas de  $h^2_a$  para todas as características avaliadas apresentaram erro padrão igual ou inferior a 0,05. Para as características diâmetro do caule (DCa), número de nós reprodutivos (NNR), uniformidade de maturação dos frutos (UMat) comprimento de folha (CF) e largura de folha (LF) foram observadas estimativas de  $h^2_a$  de magnitude moderada, no entanto, para tais características,

as estimativas de  $r_{gy}$  foram mais baixas. Para características que apresentem baixas estimativas de herdabilidades é esperado menores estimativas de capacidade preditivas (Legarra et al. 2008).

O coeficiente de correlação ou capacidade preditiva ( $r_{gy}$ ) e o coeficiente de regressão ou viés da predição ( $b$ ), envolvendo os valores fenotípicos observados e os valores genéticos preditos, são medidas práticas da capacidade dos métodos fazerem predições acuradas e não viesada, respectivamente (Resende et al. 2014a). As estimativas da capacidade preditiva ( $r_{gy}$ ) das 18 características em análises variam de -0,01 a 0,40 para as características número de nós reprodutivos (NNR) e diâmetro da copa (DCo), respectivamente (Tabela 3). Tais estimativas apresentaram erro padrão variando de 0,17 a 0,34. Além do DCo, as maiores estimativas da capacidade preditiva foram observadas nas características número de nós vegetativos (NNV) e altura de plantas (API). Os resultados de  $h^2_a$  e  $r_{gy}$  demonstram alta associação positiva, apresentando coeficiente de correlação de 88%. Assim como observado nesse trabalho, associação entre a capacidade preditiva e a herdabilidade tem sido relatada (Cavalcanti et al. 2012; Gois et al. 2016). As estimativas do viés da predição variaram de 0,25 a 1,92 para as características número de nós reprodutivos (NNR) e número de nós vegetativos (NNV). A maioria das características avaliadas apresentaram estimativa de  $b$  próxima à unidade indicando predição não viesada para as características. Portanto, são efetivas em prever as reais magnitudes das diferenças entre os indivíduos (Resende et al. 2012c). Tais estimativas apresentaram erro padrão variando de 0,77 a 3,25.

A acurácia seletiva ( $r_{gg}$ ) refere-se à correlação entre o valor genotípico verdadeiro do tratamento genético e aquele estimado ou predito a partir das informações fenotípicas (Gois et al. 2016). Valores adequados de  $r_{gg}$  são aqueles próximos à unidade. A acurácia é tanto mais alta quanto menores forem os desvios absolutos entre os valores genéticos paramétricos e os valores genéticos estimados ou preditos (Resende and Duarte 2007). O valor desta

medida indica o quão preciso é o modelo em estimar o GEBV. As estimativas da acurácia seletiva obtidas com a utilização da GWS ( $r_{gg}$ ) estão apresentadas na tabela 3. Para a característica número de nós reprodutivos (NNR) a  $r_{gg}$  não foi estimada, uma vez que para essa característica a capacidade preditiva foi próxima de zero. Para as demais características, os valores estimados de  $r_{gg}$  variaram de 0,06 para uniformidade de maturação dos frutos (UMat) e largura de folha (LF) a 0,61 para o diâmetro da copa (DCo), sendo consideradas de magnitudes baixa à moderada (Resende and Duarte 2007). Foi verificado alta correlação entre as estimativas de  $r_{gg}$  com  $h^2_a$  (82%) e com  $r_{gy}$  (99%). Correlação positiva entre acurácia seletiva e herdabilidade também foi relatada para ferrugem amarela e ferrugem do colmo, em trigo (Ornella et al. 2012).

As baixas magnitudes de  $r_{gg}$  observadas em algumas características avaliadas podem ser explicadas pelo reduzido tamanho populacional e, sobretudo, ao tamanho efetivo da população. Em estudos com populações de trigo, o aumento do tamanho populacional possibilitou incrementos nas estimativa de acurácias seletivas (Heffner et al. 2011a; Heffner et al. 2011b). O sucesso da seleção genômica é influenciada por diversos fatores, os quais interferem na acurácia seletiva de um modelo de GWS (Grattapaglia and Resende 2011; Desta and Ortiz 2014). Esses autores destacaram o tamanho da população de treinamento, tamanho efetivo populacional, densidade de marcadores, herdabilidade da característica, número de QTL que governam as características. Destes, a herdabilidade e o número de QTL envolvidos no caráter são características inerentes à arquitetura genética da característica (Resende et al. 2014a). Além disso, a estrutura genética da população pode influenciar as predições genômicas (Zhang et al. 2010; Wang et al. 2014; Li et al. 2014). Nesse sentido, as diferentes frequências alélicas entre subpopulações podem produzir associações falsas entre os dados moleculares e fenotípicos (Price et al. 2010) e, dessa forma, superestimar as herdabilidade e reduzir a acurácia seletiva (Riedelsheimer et al. 2012; Wray et al. 2013).

As estimativas dos números de QTL que governam as características ( $n_{QTL}$ ) variaram de 149 a 17.758 QTL para as características diâmetro da copa (DCo) e largura da folha (LF), respectivamente (Tabela 3). Foi observado elevado número de QTL controlando as características agronômicas, sendo estimado para as características produção de grãos de café e incidência à ferrugem, consideradas as principais características dos programas de melhoramento, 751 e 221 QTL, respectivamente. Os resultados evidenciaram relação inversamente proporcional entre a acurácia seletiva ( $r_{gg}$ ) e o número de QTL. Isso pode ser justificado pelo aumento da complexidade preditiva em função do maior do número de genes que governam a característica. Quando muitos genes estão envolvidos na manifestação de uma característica, em geral, tais genes são de efeito menor e, conseqüentemente, a estimação precisa de seus efeitos é dificultada (Goddard 2009). Isso evidencia a importância da utilização de alta densidade de marcadores SNP nas análises preditivas. Em estudos com espécies florestais (Grattapaglia and Resende 2011; Iwata et al. 2011) e em milho (Riedelsheimer et al. 2012), não foi possível verificar algum tipo de relação entre número de QTL e acurácia fenotípica ou genotípica.

A estimativa do número de indivíduos necessários ( $N_i$ ) para se obter acurácia seletiva desejada ( $r_{ggd}$ ) está apresentada na tabela 3. Observa-se a necessidade de se avaliar maior número de indivíduos quando se almeja maiores estimativas de  $r_{ggd}$ . Para se obter estimativa de acurácia seletiva de 0,7, considerada de alta magnitude (Resende and Duarte 2007), os resultados demonstraram a necessidade de se avaliar de 322 a 53.701 indivíduos para as características diâmetro da copa (DCo) e largura da folha (LF), respectivamente. Para a maioria das características foram estimados a necessidade de se avaliar mais de 1.000 indivíduos para se obter  $r_{ggd}$  igual a 0,7.

Os resultados das análises preditivas utilizando diferentes densidades de marcadores evidenciaram, de maneira geral, o aumento da acurácia seletiva ( $r_{gg}$ ) ao se elevar o número

de SNP nas análises (Tabela 4). Contudo, ao se atingir o número ótimo de marcadores, o qual maximiza as estimativas de  $r_{gg}$ , as acurácias seletivas reduziram com o aumento do número de marcadores. Assim como observado em outros estudos, o aumento do número de marcadores não aumenta linearmente a acurácia da GWS (Fernando et al. 2007; Cavalcanti et al. 2012). Esses resultados demonstraram a necessidade de avaliar outros genótipos de cafeeiros, de forma a aumentar as acurácias seletivas das características em estudo.

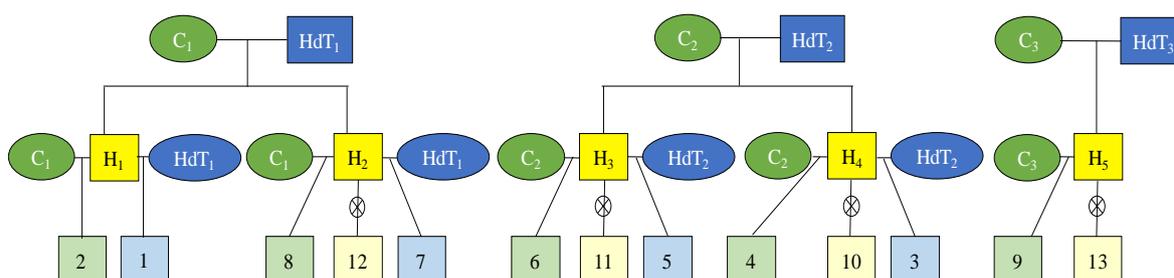
A eficiência das análises de GWS em relação à seleção fenotípica está apresentada na tabela 4. Essa análise não foi realizada para a característica número de nós reprodutivos (NNR) porque a estimativa da capacidade preditiva dessa característica foi próxima de zero (Tabela 3). Os resultados demonstraram que a possibilidade de redução do tempo do ciclo seletivo permite aos melhoristas obtenção de ganhos genéticos precocemente, sendo esta uma das grandes vantagens da GWS. Ao se reduzir o ciclo seletivo de 24 para doze anos (50%), observa-se que a GWS foi mais eficiente que a seleção com base nos dados fenotípicos para nove características avaliadas, as quais incluem incidências à ferrugem (Fer), à cercospora (Cer) e ao bicho mineiro (BM). Assim, ao se selecionar precocemente os indivíduos superiores por meio as análises de GWS, o melhorista concentrará esforços em genótipos potenciais, sendo os indesejáveis eliminados. Por essa razão, os custos de manutenção de populações de melhoramento em campo podem ser reduzidos. Além disso, a seleção genômica precoce possibilita condução de populações de melhoramento com maior desempenho agrônômico o que resultará na maximização dos ganhos genéticos. O fato da seleção com base nos dados fenotípicos ser mais eficiente que a seleção genômica para as algumas características podem ser explicados pelo número de genótipos avaliados. Outros genótipos de cafeeiro podem ser analisados com os SNPs selecionados nesse presente estudos, de forma a aumentar as acurarias seletivas.

### **Perspectivas com a utilização da GWS em *C. arabica***

Com a globalização e o expressivo aumento da população mundial, é intensificada a demanda por técnicas capazes de auxiliarem os melhoristas no desenvolvimento de novas cultivares. Nesse sentido, a elucidação e utilização da genômica permite, de forma robusta e precisa, o acesso às informações genéticas, as quais são potencialmente úteis aos programas de melhoramento do cafeeiro. O melhor entendimento da variação genética presente nas populações de melhoramento possibilitará a redução de tempo e recurso destinados ao desenvolvimento de novas cultivares. Além disso, possibilitará a seleção de linhagens/cultivares mais produtivas, adaptadas e com qualidade superior de bebida. Contudo, novos estudos são necessários, de forma a aumentar o número de genótipos que compõem as populações de estimação, validação e seleção. Além disso, abordar a utilização de diferentes métodos de GWS. Nesse sentido, avaliar a eficiência dos métodos de redes neurais artificiais, regressão quantílica bayesiana LASSO adaptativo (Alhamzawi et al. 2012), Bayes A (Meuwissen et al. 2001) e avaliar a influência da interação genótipo e ambiente. Finalmente, estudos de associação genômica (GWAS) são necessários, os quais possibilitarão a seleção de marcadores associados a genes específicos e, conseqüentemente, a seleção de genótipos superiores.

## **Conclusão**

A seleção genômica ampla mostrou-se potencial para o melhoramento de *C. arabica* por possibilitar a obtenção acurada dos GEBV e pela maior eficiência em relação a seleção fenotípica possibilitando a redução no tempo necessário para completar o ciclo de seleção.



**Figura 1.** Heredograma das 13 progênies de *Coffea arabica* oriundas de cruzamentos entre genitores do grupo de cultivares Catuaí e Híbrido de Timor (HdT);  $C_1$ ,  $C_2$  e  $C_3$  = genótipos Catuaí amarelo IAC 30, IAC 86 e IAC 64, respectivamente;  $HdT_1$ ,  $HdT_2$  e  $HdT_3$  = genótipos Híbrido de Timor UFV 445-46, UFV 440-10 e UFV 530, respectivamente;  $H_1$ ,  $H_2$ ,  $H_3$ ,  $H_4$  e  $H_5$  = híbridos oriundos de cruzamentos entre genitores Catuaí amarelos e Híbrido de Timor; 1, 3, 5 e 7 = progênies de geração de retrocruzamento resistente; 2, 4, 6, 8 e 9 = progênies de geração de retrocruzamento suscetível; 10, 11, 12 e 13 = progênies em geração  $F_2$ .

**Tabela 1.** Progenies de Coffea arabica avaliadas em Viçosa (MG).

<b>Progenies</b>	<b>Genitor 1</b>	<b>Genitor 2</b>
RCr 1	H 419-1 c-17	UFV 445-46
RCs 2	H 419-1 c-17	UFV 2143-235
RCr 3	H 514-8 c-387	UFV 440-10
RCs 4	H 514-8 c-387	UFV 2154-344
RCr 5	H 514-7 c-364	UFV 440-10
RCs 6	H 514-7 c-364	UFV 2154-344
RCr 7	H 419-10 c-214	UFV 445-46
RCs 8	H 419-10 c-214	UFV 2143-235
RCs 9	UFV 2148-57	H 513-5 c-14
F <sub>2</sub> 10	H 514-8 c-387	-
F <sub>2</sub> 11	H 514-7 c-364	-
F <sub>2</sub> 12	H 419-10 c-214	-
F <sub>2</sub> 13	H 513-5 c-14	-

RCr = Retrocruzamento resistente; RCs = Retrocruzamento suscetível; F<sub>2</sub> = geração obtida pela autofecundação dos híbridos F<sub>1</sub>.

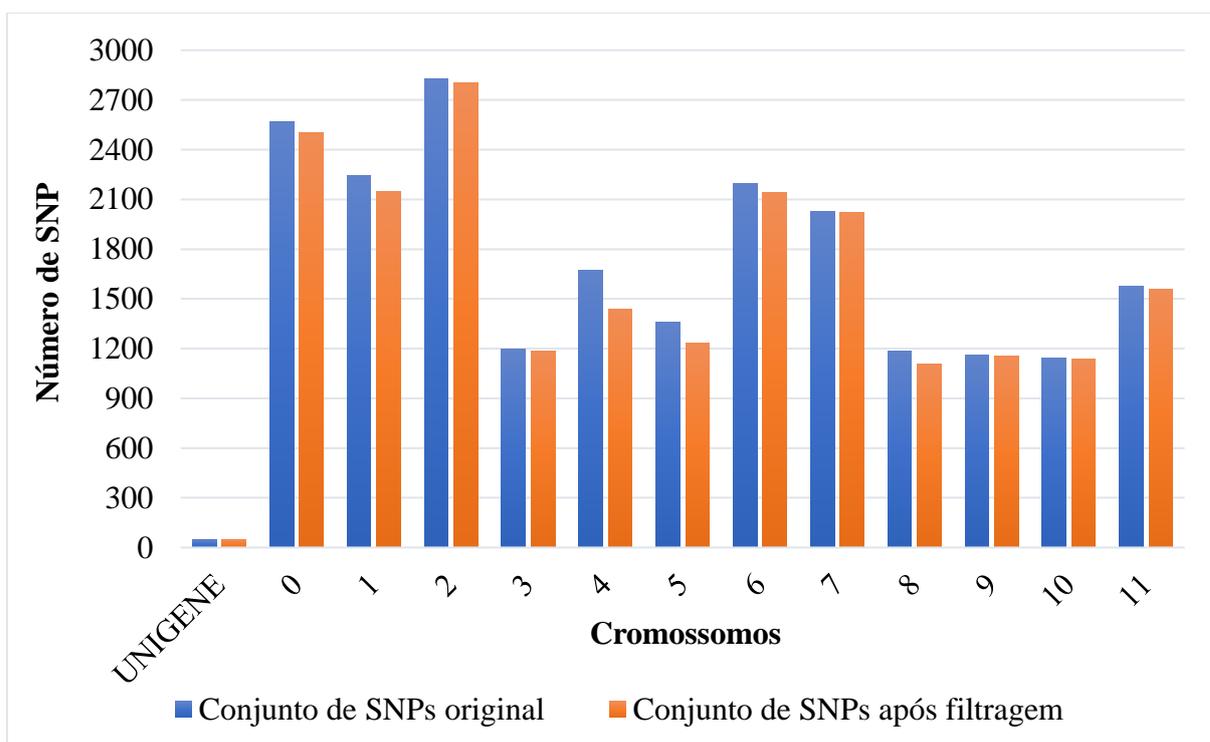
**Tabela 2.** Características fenotípicas avaliadas nos anos de 2014, 2015 e 2016 em Viçosa (MG).

<b>Característica</b>	<b>Forma de avaliação</b>
Produção	Litros de café recém colhido por planta
Comprimento de uma folha (cm)	Medidos em folha do 3º ou 4º par de um ramo plagiotrópicos do terço médio da planta
Largura de uma folha (cm)	
Comprimento de um ramo (cm)	Medidos no ramo plagiotrópico localizado no terço médio da planta
Número de nós reprodutivos	
Número de nós vegetativos	
Número total de frutos	
Volume de frutos	
Altura de plantas (cm)	Medida do ramo ortotrópico (superfície do solo até o ponto final de crescimento do ramo)
Diâmetro da copa (cm)	Medido no sentido transversal à linha de plantio, medindo-se a maior projeção da "saia" do cafeeiro
Diâmetro do caule (cm)	Medido na região do coleto da planta (cerca de 5cm da superfície do solo)
Tamanho dos frutos maduros	Avaliado por notas de 1 a 3
Uniformidade de maturação dos frutos	Avaliada por nota de 1 a 4
Ciclo de maturação dos frutos	
Incidência de Ferrugem	Avaliada por notas de 1 a 5
Incidência de Cercosporiose	
Infestação de Bicho-Mineiro	
Vigor vegetativo	Avaliado por notas de 1 a 10

**Tabela 3.** Estimativa de parâmetros genéticos obtidos por meio de análises de modelos mistos (REML/BLUP), resultados da análise da Seleção Genômica Ampla (GWS) e estimativa do número de indivíduos a serem avaliados para se obter determinada acurácia seletiva para dezoito características morfoagronômicas em população de melhoramento de *Coffea arabica*.

Característica	REML/BLUP		GWS								Ni				
	$h^2_{fen}$	$r_{yy}$	$h^2_a$	$sd_h$	$r_{gy}$	$sd_r$	<b>b</b>	$sd_b$	$r_{gg}$	nQTL	$r_{ggd\ 0,5}$	$r_{ggd\ 0,6}$	$r_{ggd\ 0,7}$	$r_{ggd\ 0,8}$	$r_{ggd\ 0,9}$
Prod	0,55	0,74	0,26	0,03	0,13	0,27	1,63	3,01	0,25	751	964	1626	2778	5140	12326
Fer	0,61	0,78	0,31	0,04	0,26	0,22	1,50	1,34	0,46	221	237	400	684	1265	3033
Cer	0,38	0,62	0,44	0,05	0,31	0,30	1,45	1,55	0,47	304	231	390	666	1233	2957
BM	0,30	0,55	0,30	0,04	0,18	0,24	1,34	1,71	0,33	476	536	904	1544	2858	6852
Vig	0,70	0,84	0,34	0,04	0,21	0,19	1,38	1,39	0,36	440	437	738	1260	2332	5592
CMat	0,72	0,85	0,31	0,05	0,12	0,19	1,31	2,15	0,21	1313	1434	2421	4134	7650	18345
UMat	0,30	0,55	0,28	0,03	0,03	0,27	0,62	3,13	0,06	14775	17841	30107	51424	95152	228177
TFr	0,50	0,71	0,36	0,04	0,23	0,24	1,52	2,01	0,39	394	370	624	1066	1973	4730
CF	0,42	0,65	0,29	0,02	0,06	0,23	0,87	2,21	0,12	3981	4530	7644	13057	24160	57935
LF	0,44	0,66	0,32	0,05	0,03	0,25	0,44	2,62	0,06	17758	18631	31440	53701	99365	238281
CRP	0,78	0,88	0,41	0,04	0,32	0,34	1,20	1,30	0,50	244	198	335	572	1058	2538
NNR	0,49	0,70	0,23	0,02	-0,01	0,21	0,25	3,25	-	-	-	-	-	-	-
NNV	0,44	0,66	0,46	0,04	0,38	0,21	1,92	1,63	0,56	199	143	242	413	765	1834
NFRP	0,49	0,70	0,34	0,05	0,14	0,20	1,33	2,10	0,23	1157	1134	1913	3267	6046	14498
VFRP	0,57	0,76	0,25	0,03	0,11	0,17	1,23	1,90	0,21	1081	1418	2393	4087	7562	18133
API	0,90	0,95	0,46	0,04	0,38	0,18	1,18	0,77	0,56	202	146	246	420	777	1864
DCo	0,90	0,95	0,45	0,03	0,40	0,22	1,46	0,90	0,61	149	112	189	322	596	1429
DCa	0,01	0,10	0,16	0,01	0,06	0,26	1,14	2,90	0,14	1658	3363	5674	9692	17934	43006

$h^2_{fen}$  = Herdabilidade fenotípica;  $r_{yy}$  = Acurácia da seleção obtida pelo método REML/BLUP utilizando dados fenótipos;  $h^2_a$  = herdabilidade genômica;  $sd_h$  = erro padrão das estimativas de  $h^2_a$ ;  $r_{gy}$  = Capacidade preditiva da GWS;  $sd_r$  = erro padrão das estimativas de  $r_{gy}$ ; **b** = viés da predição;  $sd_b$  = erro padrão das estimativas de **b**;  $r_{gg}$  = Acurácia seletiva da GWS; nQTL = Estimativa do número de QTL que controlam a característica;  $r_{ggd}$  = acurácia seletiva desejada; Ni = número de indivíduos avaliados para se obter determinada  $r_{ggd}$ ; Prod = Produção de frutos de café; Fer = Incidência de ferrugem; Cer = Incidência de Cercospora; BM = Infestação de Bicho-Mineiro; Vig = Vigor vegetativo; CMat = Ciclo de maturação dos frutos; UMat = Uniformidade de maturação dos frutos; TFr = Tamanho dos frutos maduros; CF = Comprimento de uma folha; LF = Largura de uma folha; CRP = Comprimento de um ramo plagiotrópico; NNR = Número de nós reprodutivos; NNV = Número de nós vegetativos; NFRP = Número de frutos por ramo plagiotrópico; VFRP = Volume de frutos por ramo plagiotrópico; API = Altura de planta; DCo = Diâmetro da copa; DCa = Diâmetro do caule.



**Figura 2.** Marcadores moleculares SNP distribuídos ao longo dos UNIGENES proveniente das sequências de ESTs de *Coffea arabica* e dos 11 cromossomos e do “cromossomo 0” de *Coffea canephora*. “Cromossomo 0” é apenas um conjunto de scaffolds de sequência não ordenada (Denoeud et al. 2014).

**Tabela 4.** Acurácia seletiva estimada a partir de diferentes densidades de marcadores SNP e eficiência da seleção genômica ampla (GWS) em relação à seleção fenotípica em população de melhoramento de *Coffea arabica*.

Característica	<sup>a</sup> Número de SNP							<sup>b</sup> Anos para análise de GWS	
	1.000	4.000	8.000	12.000	16.000	20.000	20.477	24	12
Prod	-0,08	0,10	0,21	0,25	0,27	0,26	0,25	0,34	0,68
Fer	0,19	0,42	0,40	0,24	0,24	0,38	0,46	0,59	1,18
Cer	0,39	0,37	0,35	0,36	0,41	0,44	0,47	0,76	1,52
BM	0,23	0,28	0,23	0,23	0,27	0,26	0,33	0,60	1,20
Vig	0,27	0,46	0,43	0,37	0,49	0,37	0,36	0,43	0,86
CMat	-0,07	0,11	0,01	0,27	0,10	0,18	0,21	0,25	0,49
UMat	0,07	0,15	0,17	0,09	0,01	0,25	0,06	0,11	0,22
TFr	0,33	0,47	0,30	0,40	0,31	0,36	0,39	0,54	1,09
CF	0,14	0,18	0,22	0,07	0,12	0,16	0,12	0,18	0,37
LF	0,00	0,26	0,11	0,27	0,22	0,19	0,06	0,09	0,18
CRP	0,42	0,44	0,50	0,51	0,51	0,56	0,50	0,56	1,13
NNR	0,03	0,04	-0,07	0,12	0,09	0,06	-0,01	-	-
NNV	0,44	0,33	0,43	0,44	0,46	0,50	0,56	0,84	1,68
NFRP	0,08	0,18	0,29	0,33	0,27	0,25	0,23	0,33	0,67
VFRP	0,23	0,27	0,30	0,24	0,21	0,19	0,21	0,28	0,55
API	0,56	0,48	0,58	0,57	0,48	0,54	0,56	0,59	1,17
DCo	0,37	0,48	0,52	0,58	0,59	0,57	0,61	0,64	1,28
DCa	-0,03	0,14	0,16	-0,04	-0,02	-0,05	0,14	1,38	2,77

<sup>a</sup> = Acurácia seletiva estimada a partir de diferentes densidades de marcadores SNP; <sup>b</sup> = Eficiência da seleção genômica ampla (GWS) em relação à seleção fenotípica; Prod = Produção de frutos de café; Fer = Incidência de ferrugem; Cer = Incidência de Cercospora; BM = Infestação de Bicho-Mineiro; Vig = Vigor vegetativo; CMat = Ciclo de maturação dos frutos; UMat = Uniformidade de maturação dos frutos; TFr = Tamanho dos frutos maduros; CF = Comprimento de uma folha; LF = Largura de uma folha; CRP = Comprimento de um ramo plagiotrópico; NNR = Número de nós reprodutivos; NNV = Número de nós vegetativos; NFRP = Número de frutos por ramo plagiotrópico; VFRP = Volume de frutos por ramo plagiotrópico; API = Altura de planta; DCo = Diâmetro da copa; DCa = Diâmetro do caule.

## Referências Bibliográficas

- Alhamzawi R, Yu K, Benoit DF (2012) Bayesian adaptive Lasso quantile regression. *Statistical Modelling* 12:279–297. doi: 10.1177/1471082X1101200304
- Alkimim ER, Caixeta ET, Sousa TV, et al (2017) Marker-assisted selection provides arabica coffee with genes from other *Coffea* species targeting on multiple resistance to rust and coffee berry disease. *Molecular Breeding* 37:6. doi: 10.1007/s11032-016-0609-1
- Barbosa MHP, Resende MDV, Dias LA dos S, et al (2012) Genetic improvement of sugar cane for bioenergy: the brazilian experience in network research with RIDESA. *Crop Breeding and Applied Biotechnology* 12:87–98. doi: 10.1590/S1984-70332012000500010
- Bernardo R (2008) Molecular Markers and Selection for Complex Traits in Plants: Learning from the Last 20 Years. *Crop Science* 48:1649. doi: 10.2135/cropsci2008.03.0131
- Bernardo R, Yu J (2007) Prospects for Genomewide Selection for Quantitative Traits in Maize. *Crop Science* 47:1082. doi: 10.2135/cropsci2006.11.0690
- Bhering LL (2017) Rbio: A tool for biometric and statistical analysis using the R platform. *Crop Breeding and Applied Biotechnology* 17:187–190. doi: 10.1590/1984-70332017v17n2s29
- Cavalcanti JJV, Resende MDV de, Santos FHC dos, Pinheiro CR (2012) Predição simultânea dos efeitos de marcadores moleculares e seleção genômica ampla em cajueiro. *Revista Brasileira de Fruticultura* 34:840–846. doi: 10.1590/S0100-29452012000300025
- Cenci A, Combes M-C, Lashermes P (2012) Genome evolution in diploid and tetraploid *Coffea* species as revealed by comparative analysis of orthologous genome segments. *Plant molecular biology* 78:135–145. doi: 10.1007/s11103-011-9852-3
- Clarindo WR, Carvalho CR (2008) First *Coffea arabica* karyogram showing that this species

- is a true allotetraploid. *Plant Systematics and Evolution* 274:237–241. doi: 10.1007/s00606-008-0050-y
- Corrêa TR, Motoike SY, Coser SM, et al (2015) Estimation of genetic parameters for in vitro oil palm characteristics (*Elaeis guineensis* Jacq.) and selection of genotypes for cloning capacity and oil yield. *Industrial Crops and Products* 77:1033–1038. doi: 10.1016/j.indcrop.2015.09.066
- Crossa J, Campos G d. I., Perez P, et al (2010) Prediction of Genetic Values of Quantitative Traits in Plant Breeding Using Pedigree and Molecular Markers. *Genetics* 186:713–724. doi: 10.1534/genetics.110.118521
- Davey JW, Hohenlohe PA, Etter PD, et al (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics* 12:499–510. doi: 10.1038/nrg3012
- De Los Campos G, Hickey JM, Pong-Wong R, et al (2013) Whole-Genome Regression and Prediction Methods Applied to Plant and Animal Breeding. *Genetics* 193:327–345. doi: 10.1534/genetics.112.143313
- Denoeud F, Carretero-Paulet L, Dereeper A, et al (2014) The coffee genome provides insight into the convergent evolution of caffeine biosynthesis. *Science (New York, NY)* 345:1181–1184. doi: 10.1126/science.1255274
- Desta ZA, Ortiz R (2014) Genomic selection: genome-wide prediction in plant improvement. *Trends in Plant Science* 19:592–601. doi: 10.1016/j.tplants.2014.05.006
- Diniz LEC, Sakiyama NS, Lashermes P, et al (2005) Analysis of AFLP markers associated to the Mex-1 resistance locus in Icatu progenies. *Cropps Breeding and Applied Biotechnology* 5:387–393. doi: 10.12702/1984-7033.v05n04a03
- Favoretto P, da Silva CC, Tavares AG, et al (2017) Assisted-selection of naturally caffeine-free coffee cultivars—characterization of SNPs from a methyltransferase gene.

- Molecular Breeding 37:31. doi: 10.1007/s11032-017-0636-6
- Fernando RL, Habier D, Stricker C, et al (2007) Genomic selection. *Acta Agriculturae Scandinavica, Section A - Animal Science* 57:192–195. doi: 10.1080/09064700801959395
- Ferreira RT, Viana AP, Barroso DG, et al (2012) Toona ciliata genotype selection with the use of individual BLUP with repeated measures. *Scientia Agricola* 69:210–216. doi: 10.1590/S0103-90162012000300006
- Ganal MW, Altmann T, Röder MS (2009) SNP identification in crop plants. *Current Opinion in Plant Biology* 12:211–217. doi: 10.1016/j.pbi.2008.12.009
- Garcia C, Lima B, Almeida A, et al (2011) Genome wide selection for Eucalyptus improvement at international paper in Brazil. *BMC Proceedings* 5:44. doi: 10.1186/1753-6561-5-S7-P44
- Gianola D (2006) Genomic-Assisted Prediction of Genetic Value With Semiparametric Procedures. *Genetics* 173:1761–1776. doi: 10.1534/genetics.105.049510
- Gianola D (2013) Priors in Whole-Genome Regression: The Bayesian Alphabet Returns. *Genetics* 194:573–596. doi: 10.1534/genetics.113.151753
- Goddard M (2009) Genomic selection: prediction of accuracy and maximisation of long term response. *Genetica* 136:245–257. doi: 10.1007/s10709-008-9308-0
- Gois IB, Borém A, Cristofani-Yaly M, et al (2016) Genome wide selection in Citrus breeding. *Genetics and Molecular Research*. doi: 10.4238/gmr15048863
- Grattapaglia D, Resende MD V. (2011) Genomic selection in forest tree breeding. *Tree Genetics & Genomes* 7:241–255. doi: 10.1007/s11295-010-0328-4
- Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME (2009) Invited review: Genomic selection in dairy cattle: Progress and challenges. *Journal of Dairy Science* 92:433–443. doi: 10.3168/jds.2008-1646

- Heffner EL, Jannink J-L, Iwata H, et al (2011a) Genomic Selection Accuracy for Grain Quality Traits in Biparental Wheat Populations. *Crop Science* 51:2597. doi: 10.2135/cropsci2011.05.0253
- Heffner EL, Jannink J-L, Sorrells ME (2011b) Genomic Selection Accuracy using Multifamily Prediction Models in a Wheat Breeding Program. *The Plant Genome Journal* 4:65. doi: 10.3835/plantgenome.2010.12.0029
- Heffner EL, Lorenz AJ, Jannink J-L, Sorrells ME (2010) Plant Breeding with Genomic Selection: Gain per Unit Time and Cost. *Crop Science* 50:1681. doi: 10.2135/cropsci2009.11.0662
- Heffner EL, Sorrells ME, Jannink J-L (2009) Genomic Selection for Crop Improvement. *Crop Science* 49:1. doi: 10.2135/cropsci2008.08.0512
- Henderson CR (1975) Best linear unbiased estimation and prediction under a selection model. *Biometrics* 31:423–449.
- Heslot N, Jannink J-L, Sorrells ME (2015) Perspectives for Genomic Selection Applications and Research in Plants. *Crop Science* 55:1. doi: 10.2135/cropsci2014.03.0249
- Iwata H, Hayashi T, Tsumura Y (2011) Prospects for genomic selection in conifer breeding: a simulation study of *Cryptomeria japonica*. *Tree Genetics & Genomes* 7:747–758. doi: 10.1007/s11295-011-0371-9
- Kwok P-Y, Gu Z (1999) Single nucleotide polymorphism libraries: why and how are we building them? *Molecular Medicine Today* 5:538–543. doi: 10.1016/S1357-4310(99)01601-9
- Lande R, Thompson R (1990) Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics* 124:743–56.
- Lashermes P, Combes M-C, Robert J, et al (1999) Molecular characterisation and origin of the *Coffea arabica* L. genome. *Molecular and General Genetics MGG* 261:259–266.

doi: 10.1007/s004380050965

- Lashermes P, Hueber Y, Combes M-C, et al (2016) Inter-genomic DNA Exchanges and Homeologous Gene Silencing Shaped the Nascent Allopolyploid Coffee Genome (*Coffea arabica* L.). *Genes|Genomes|Genetics* 6:2937–2948. doi: 10.1534/g3.116.030858
- Legarra A, Robert-Granie C, Manfredi E, Elsen J-M (2008) Performance of Genomic Selection in Mice. *Genetics* 180:611–618. doi: 10.1534/genetics.108.088575
- Li M, Liu X, Bradbury P, et al (2014) Enrichment of statistical power for genome-wide association studies. *BMC Biology* 12:73. doi: 10.1186/s12915-014-0073-5
- Longin CFH, Mi X, Würschum T (2015) Genomic selection in wheat: optimum allocation of test resources and comparison of breeding strategies for line and hybrid breeding. *Theoretical and Applied Genetics* 128:1297–1306. doi: 10.1007/s00122-015-2505-1
- Meuwissen TH, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157:1819–1829. doi: 11290733
- Ornella L, Singh S, Perez P, et al (2012) Genomic Prediction of Genetic Values for Resistance to Wheat Rusts. *The Plant Genome Journal* 5:136. doi: 10.3835/plantgenome2012.07.0017
- Patel DA, Zander M, Dalton-Morgan J (2015) Advances in Plant Genotyping: Where the Future Will Take Us. In: Batley J (ed) *Plant genotyping*. Springer, New York, pp 1–11
- Patterson HD, Thompson R (1971) Recovery of inter-block information when block sizes are unequal. *Biometrika* 58:545–554. doi: 10.1093/biomet/58.3.545
- Pereira TB, Carvalho JPF, Botelho CE, et al (2013) Eficiência da seleção de progênies de café F4 pela metodologia de modelos mistos (REML/BLUP). *Bragantia* 72:230–236. doi: 10.1590/brag.2013.031
- Perez P, de los Campos G (2014) Genome-Wide Regression and Prediction with the BGLR

- Statistical Package. *Genetics* 198:483–495. doi: 10.1534/genetics.114.164442
- Poland J (2015) Breeding-assisted genomics. *Current Opinion in Plant Biology* 24:119–124. doi: 10.1016/j.pbi.2015.02.009
- Price AL, Zaitlen NA, Reich D, Patterson N (2010) New approaches to population stratification in genome-wide association studies. *Nature Reviews Genetics* 11:459–463. doi: 10.1038/nrg2813
- R Core Team (2017) R: A Language and Environment for Statistical Computing.
- Resende M, Caixeta E, Alkimim ER, et al (2016) High-Throughput Targeted Genotyping of *Coffea arabica* and *Coffea canephora* Using Next Generation Sequencing. San Diego, CA, p 1
- Resende M, Silva F, Resende Junior, MFR Azevedo C (2014a) Genome-wide selection. In: Borem A, Fritsche-Neto R (eds) *Biotechnology and Plant Breeding*, Elsevier.
- Resende MDV, Duarte JB (2007) Precisão e controle de qualidade em experimentos de avaliação de cultivares. *Pesquisa Agropecuária Tropical* 37:182–194.
- Resende MDV, Thompson R (2004) Factor analytic multiplicative mixed models in the analysis of multiple experiments. *Revista de Matemática e Estatística* 22:1–22.
- Resende MFRJ, Muñoz P, Acosta JJ, et al (2012a) Accelerating the domestication of trees using genomic selection: accuracy of prediction models across ages and environments. *New Phytologist* 193:617–624. doi: 10.1111/j.1469-8137.2011.03895.x
- Resende MFRJ, Munoz P, Resende MD V., et al (2012b) Accuracy of Genomic Selection Methods in a Standard Data Set of Loblolly Pine (*Pinus taeda* L.). *Genetics* 190:1503–1510. doi: 10.1534/genetics.111.137026
- Resende MDV de (2016) Software Selegen-REML/BLUP: a useful tool for plant breeding. *Crop Breeding and Applied Biotechnology* 16:330–339. doi: 10.1590/1984-70332016v16n4a49

- Resende MDV de (2007) Matemática e estatística na análise de experimentos e no melhoramento genético. Colombo: Embrapa Florestas
- Resende MDV de, Lopes P, Silva RL Da, Pires IE (2008) Seleção genômica ampla (GWS) e maximização da eficiência do melhoramento genético. Pesquisa Florestal Brasileira 56:63–77.
- Resende MDV de, Silva FF e, Azevedo CF (2014b) Estatística Matemática, Biométrica e Computacional: Modelos Mistos, Multivariados, Categorias e Generalizados (REML/BLUP), Inferência Bayesiana, Regressão Aleatória, Seleção Genômica, QTI-GWAS, Estatística Espacial e Temporal, Competição, Sobrevivência, 1st edn. Suprema, Viçosa
- Resende MD V. (2008) Genômica Quantitativa e Seleção no Melhoramento de Plantas Perenes e Animais, Colombo: E.
- Resende MD V, Resende MFR, Sansaloni CP, et al (2012c) Genomic selection for growth and wood quality in Eucalyptus: capturing the missing heritability and accelerating breeding for complex traits in forest trees. New Phytologist 194:116–128. doi: 10.1111/j.1469-8137.2011.04038.x
- Resende MD V, Silva FF, Lopes PS, Azevedo CF (2012d) Seleção genômica ampla (GWS) via modelos mistos (REML/BLUP), inferência bayesiana (MCMC), regressão aleatória multivariada e estatística espacial. Viçosa
- Riedelsheimer C, Czedik-Eysenberg A, Grieder C, et al (2012) Genomic and metabolic prediction of complex heterotic traits in hybrid maize. Nature Genetics 44:217–220. doi: 10.1038/ng.1033
- Soller M, Beckmann JS (1983) Genetic polymorphism in varietal identification and genetic improvement. Theoretical and Applied Genetics 67:25–33. doi: 10.1007/BF00303917
- Spinelli VM, Dias LAS, Rocha RB, Resende MD V. (2015) Estimates of genetic parameters

- with selection within and between half-sib families of *Jatropha curcas* L. *Industrial Crops and Products* 69:355–361. doi: 10.1016/j.indcrop.2015.02.024
- Viana JMS, Faria VR, Silva FF e, Resende MDV de (2011) Best Linear Unbiased Prediction and Family Selection in Crop Species. *Crop Science* 51:2371. doi: 10.2135/cropsci2011.03.0153
- Vidal RO, Mondego JMC, Pot D, et al (2010) A High-Throughput Data Mining of Single Nucleotide Polymorphisms in Coffea Species Expressed Sequence Tags Suggests Differential Homeologous Gene Expression in the Allotetraploid *Coffea arabica*. *Plant Physiology* 154:1053–1066. doi: 10.1104/pp.110.162438
- Wang C, Zhan X, Bragg-Gresham J, et al (2014) Ancestry estimation and control of population stratification for sequence-based association studies. *Nature Genetics* 46:409–415. doi: 10.1038/ng.2924
- Wong CK, Bernardo R (2008) Genomewide selection in oil palm: increasing selection gain per unit time and cost with small populations. *Theoretical and Applied Genetics* 116:815–824. doi: 10.1007/s00122-008-0715-5
- Wray NR, Yang J, Hayes BJ, et al (2013) Pitfalls of predicting complex traits from SNPs. *Nature Reviews Genetics* 14:507–515. doi: 10.1038/nrg3457
- Xu Y, Lu Y, Xie C, et al (2012) Whole-genome strategies for marker-assisted plant breeding. *Molecular Breeding* 29:833–854. doi: 10.1007/s11032-012-9699-6
- Zhang Z, Ersoz E, Lai C-Q, et al (2010) Mixed linear model approach adapted for genome-wide association studies. *Nature Genetics* 42:355–360. doi: 10.1038/ng.546