

EMILLY RUAS ALKIMIM

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

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

A415d
2017 Alkimim, Emilly Ruas, 1989-
Diversidade genética, ganhos com seleção e seleção
genômica ampla na espécie *Coffea canephora* / Emilly Ruas
Alkimim. – Viçosa, MG, 2017.
xiv, 105f. : il. (algumas color.) ; 29 cm.

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

1. Café-Melhoramento genético. 2. Genômica.
3. Marcadores genéticos. I. Universidade Federal de Viçosa.
Outros Órgãos. Programa de Pós-graduação em Genética e
Melhoramento. II. Título.

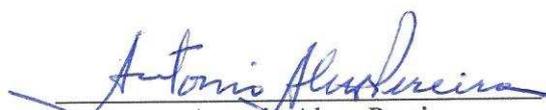
CDD 22. ed. 633.73

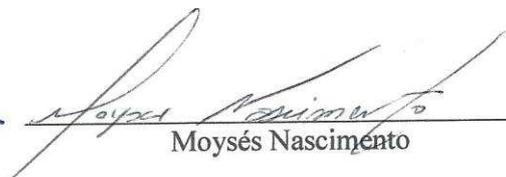
EMILLY RUAS ALKIMIM

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

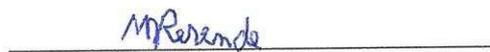
Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento, para obtenção do título de *Doctor Scientiae*.

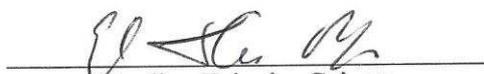
APROVADA: 31 de julho de 2017.


Antonio Alves Pereira


Moysés Nascimento


Fábio Daniel Tancredi


Marcos Deon Vilela de Resende
(Coorientador)


Eveline Teixeira Caixeta
(Orientadora)

“Em seu coração o homem planeja o seu caminho, mas o Senhor determina os seus passos”.

Provérbios

16:9

Dedico a Deus, que esteve comigo em todos os momentos guiando os meus passos. E aos meus pais e meu esposo pelo amor, carinho e compreensão.

AGRADECIMENTOS

A Deus pelo dom da vida, por sempre guiar os meus passos me dando forças para vencer os momentos difíceis e por me permitir celebrar com alegria as vitórias alcançadas. Sem Deus eu nada seria!

Agradeço a Universidade Federal De Viçosa (UFV) e ao Curso de Pós-Graduação em Genética 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, a quem posso chamar de amiga, pela ajuda, orientação e ensinamentos transmitidos ao longo de todos esses anos. Obrigada por tudo!!!

Aos meus coorientadores Dr. Marcos Deon Vilela de Resende e Dr. Felipe Lopes da Silva por toda ajuda na condução dos trabalhos e pela amizade construída. A ajuda de vocês foi de suma importância.

Aos Drs. Moysés Nascimento, Antonio Alves Pereira e Fábio Daniel Tancredi por aceitarem com alegria o convite para participar da minha banca de defesa de tese e por compartilharem seus conhecimentos.

Agradeço aos meus pais, Jodecy e Sheila, que são exemplos para mim, pelo inestimável amor, carinho e esforço dedicado e por sempre acreditarem e confiarem em mim. Sem a confiança e ajuda deles a minha caminhada com certeza não teria sido a mesma. Amo vocês!

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

Ao meu esposo Tiago, a quem amo muito, por estar sempre ao meu lado em toda essa caminhada gratificante, me ajudando, incentivando e acreditando em mim.

Aos colegas do Laboratório BioCafé, pela eterna amizade, pelos conselhos e por dividir comigo grandes momentos. Em especial agradeço aos amigos Geleta, Juan, Fran, Letícia, Dani e Júlia, por todos os momentos de alegria e também de tristezas que compartilhamos.

Aos amigos que a vida e UFV me proporcionou ao longo da minha caminhada. Muito obrigada a todos pela amizade!

BIOGRAFIA

EMILLY RUAS ALKIMIM, filha de Jodecy Soares Alkimim e Sheila Cristina Ruas Alkimim, nasceu no dia 03 de agosto de 1989, em Francisco Sá, Estado de Minas Gerais.

Concluiu o ensino fundamental em 2003, na Escola Estadual Tiburtino Pena em Francisco Sá-MG e o ensino médio em 2006, no Colégio Pirâmide em Francisco Sá-MG.

Em dezembro de 2011 diplomou-se em Agronomia na Universidade Estadual de Montes Claros (UNIMONTES). Durante a graduação foi bolsista de iniciação científica no Laboratório de Biotecnologia, 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, na Universidade Federal de Viçosa (UFV), sob orientação da Dra. Eveline Teixeira Caixeta, pesquisadora da Embrapa Café.

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, pesquisadora da Embrapa Café, submetendo-se a defesa de tese em Julho de 2017.

SUMÁRIO

LISTA DE TABELAS	ix
LISTA DE FIGURAS	xi
RESUMO	xii
ABSTRACT	xiv
Introdução geral	1
Referências	3

CAPÍTULO 1

High-throughput targeted genotyping and validation of SNP markers in <i>Coffea canephora</i> using next generation sequencing	4
ABSTRACT	5
Introduction	6
Material and methods	8
<i>Genetic material</i>	8
<i>Genomic DNA extraction and SNP identification</i>	8
<i>Quality analysis of SNPs</i>	9
<i>Analysis of genetic diversity and population structure</i>	9
Results	10
Discussion	12
Conclusions	16
References	29

CAPÍTULO 2

Ganhos com seleção e análise de diversidade genética em populações de <i>Coffea canephora</i> via modelos mistos (REML/BLUP)	34
RESUMO	35
Introdução	36
Material e métodos	37
<i>Material genético</i>	37
<i>Avaliações fenotípicas</i>	38
<i>Análise de Deviance (ANADEV)</i>	39

<i>Parâmetros genéticos dos genótipos Conilon e Robusta</i>	39
<i>Parâmetros genéticos das famílias híbridas</i>	41
<i>Correlação genética</i>	43
<i>Índice de seleção</i>	44
<i>Diversidade genética</i>	44
<i>Análises conjunta de índice de seleção</i>	44
<i>Seleção de Híbridos para as variáveis incidência à ferrugem e produção por planta</i>	44
Resultados e discussão	44
<i>Análises dos genótipos do grupo varietal Conilon</i>	44
<i>Análises dos genótipos do grupo varietal Robusta</i>	50
<i>Análises das famílias híbridas</i>	53
<i>Análises conjunta de índice de seleção</i>	57
<i>Seleção de Híbridos para as variáveis incidência à ferrugem e produção por planta</i>	57
Conclusões	58
Referências.....	81
 CAPÍTULO 3	
Seleção genômica ampla em <i>Coffea canephora</i>	84
RESUMO	85
Introdução	86
Material e métodos.....	87
<i>Material genético</i>	87
<i>Avaliações fenotípicas</i>	88
<i>Análise dos dados fenotípicos</i>	89
<i>Extração do DNA genômico, identificação e análise de qualidade de SNP</i>	89
<i>Predição utilizando o modelo GBLUP</i>	90
<i>Validação cruzada</i>	90
<i>Capacidade preditiva, viés da predição, acurácia da GWS e nQTL</i>	91
<i>Eficiência da GWS</i>	91
Resultados e discussão	91
Conclusão.....	95
Referências.....	101

Conclusões gerais..... 105

LISTA DE TABELAS

CAPÍTULO 1

Table 1. Parents Conilon and Robusta	18
Table 2. Description of the crosses to obtain the Hybrids that make up the population	19
Table 3. Genotypes of the Conilon and Robusta varietal groups	20
Table 4. Lower values of genetic dissimilarity and mean genetic dissimilarity obtained between and within the varietal groups (Conilon and Robuta) and the Hybrids.	23
Table 5. Greater values of genetic dissimilarity obtained between and within the varietal groups (Conilon and Robusta) and Hybrids.....	26
Table 6. Greater values of genetic dissimilarity obtained between and within the varietal groups (Conilon and Robusta) and the Hybrids, disregarding the accessions Conilon UFV 3628-2 and Conilon UFV 3628-45, Robusta UFV 3374-29, and the Hybrid H0915-1.	27

CAPÍTULO 2

Tabela 1. Genótipos do grupo varietal Conilon	59
Tabela 2. Genótipos do grupo varietal Robusta	60
Tabela 3. Famílias híbridas e a descrição dos cruzamentos para obtenção das mesmas	61
Tabela 4. Análise de Deviance (ANADEV) e teste da razão de verossimilhança (TRV) para as variáveis avaliadas na população de Conilon	62
Tabela 5. Estimativa dos parâmetros genéticos e ambientais para as variáveis avaliadas na população de Conilon	63
Tabela 6. Correlação genética entre as variáveis avaliadas na população de Conilon	64
Tabela 7. Classificação dos 15 genótipos Conilon superiores com base no rank médio (índice de Mulamba-Mock), referente a oito variáveis fenotípicas	65
Tabela 8. Agrupamento dos 71 genótipos Conilon pelo método de Tocher, com base na distância genética euclidiana média	66
Tabela 9. Agrupamento dos 70 genótipos Conilon pelo método de Tocher, com base na distância genética euclidiana média	67
Tabela 10. Análise de Deviance (ANADEV) e teste da razão de verossimilhança (TRV) para as variáveis avaliadas na população de Robusta	68

Tabela 11. Estimativa dos parâmetros genéticos e ambientais para as variáveis avaliadas na população de Robusta.....	69
Tabela 12. Correlação genética entre as variáveis avaliadas na população de Robusta	70
Tabela 13. Classificação dos 15 melhores genótipos Robusta com base no rank médio (índice de Mulamba-Mock), referente a sete variáveis fenotípicas	71
Tabela 14. Agrupamento dos 56 genótipos Robusta pelo método de Tocher, com base na distância genética euclidiana média	72
Tabela 15. Análise de Deviance (ANADEV) e teste da razão de verossimilhança (TRV) para as variáveis avaliadas nas famílias híbridas	73
Tabela 16. Estimativa dos parâmetros genéticos e ambientais para as variáveis avaliadas nas famílias híbridas	74
Tabela 17. Correlação genética entre as variáveis avaliadas nas famílias híbridas ..	75
Tabela 18. Classificação das famílias híbridas com base no rank médio (índice de Mulamba-Mock), referente a quatro variáveis fenotípicas	76
Tabela 19. Agrupamento das 20 famílias híbridas pelo método de Tocher, com base na distância genética euclidiana média	77
Tabela 20. Classificação das famílias híbridas e seus genitores com base no rank médio (índice de Mulamba-Mock), referente a quatro variáveis fenotípicas	78
Tabela 21. Seleção dos 20 melhores Híbridos dentro das famílias híbridas, para a variável incidência à ferrugem, com base nos efeitos aditivos (a) e valores genéticos aditivos (u + a)	79
Tabela 22. Seleção dos 20 melhores Híbridos dentro das famílias híbridas, para a variável produção por planta, com base nos efeitos aditivos (a) e valores genéticos aditivos (u + a)	80

CAPÍTULO 3

Tabela 1. Genótipos do grupo varietal Conilon	96
Tabela 2. Híbridos e a descrição dos cruzamentos para obtenção dos mesmos.....	97
Tabela 3. Resultados da Seleção Genômica Ampla (GWS) para oito característica morfoagronômicas em população de melhoramento de <i>Coffea canephora</i>	98
Tabela 4. Número de indivíduos necessários para se obter acurácia desejada da GWS (r_{ggd}) em população de melhoramento de <i>Coffea canephora</i> para as características Vig, Fer, Cer, AP e DC	99

LISTA DE FIGURAS

CAPÍTULO 1

- Fig 1** Dendrogram obtained by the UPGMA technique, based on the arithmetic complement of the unweighted index of 72 *Coffea canephora* genotypes (26 Conilons, 26 Robustas, and 20 Hybrids) and 33,485 SNP markers. Green, red and blue lines represent the varietal groups (Conilon and Robusta) and the Hybrids, respectively. 21
- Fig 2** Neighbor Joining Dendrogram (NJ) of the 72 *Coffea canephora* genotypes (26 Conilons, 26 Robustas and 20 Hybrids). Green, red and blue lines represent the varietal groups (Conilon and Robusta) and the Hybrids, respectively..... 22
- Fig 3** Graphic obtained with the values of ΔK for visualization of the best K ($K = 2$), according to the methodology proposed by Evanno et al. (2005)..... 23
- Fig 4** Bar graphic of the STRUCTURE software used to study the diversity of the 72 *Coffea canephora* genotypes (26 Conilons, 26 Robustas, and 20 Hybrids). The 72 genotypes are represented below the graphic, and were divided into two ($k = 2$) and three ($k = 3$) groups, Figures 5a and b respectively, according to the Conilon and Robusta varietal group, and also allowed discriminating Hybrid genotypes resulting from crosses between the two heterotic groups. The red bar is associated to the Robusta genotypes, the green bar to the Conilon genotypes and the blue bar to the Hybrid genotypes 24

CAPÍTULO 3

- Figura 1.** Eficiência da GWS em relação a seleção baseada somente em dados fenotípicos em população de melhoramento de *Coffea canephora*. 100

RESUMO

ALKIMIM, Emilly Ruas, D.Sc., Universidade Federal de Viçosa, julho de 2017. **Diversidade genética, ganhos com seleção e seleção genômica ampla na espécie *Coffea canephora*.** Orientadora: Eveline Teixeira Caixeta. Coorientadores: Marcos Deon Vilela de Resende e Felipe Lopes da Silva.

A presença de variabilidade genética associada a estratégias mais eficientes de seleção são imprescindíveis para se obter sucesso nos programas de melhoramento. Nesse sentido, o uso de marcadores moleculares em estudos de diversidade, bem como a seleção baseada em dados fenotípicos por meio do uso de procedimentos genético-estatísticos mais refinados têm se mostrado ferramentas úteis nos programas de melhoramento. Além disso, com o avanço das plataformas de NGS (*Next Generation Sequencing*), um novo método de seleção, que visa utilizar dados fenótipos e moleculares, denominado Seleção Genômica Ampla foi proposto. O objetivo do trabalho foi identificar marcadores SNP (*Single Nucleotide Polymorphism*) 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 (*Genome Wide Selection*) e avaliar sua eficiência em população de *C. canephora*. 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 192 genótipos de cafeeiros. Por fim, a população avaliada por meio da GWS foi composta por 165 genótipos. Utilizando a plataforma de sequenciamento da empresa RAPiD Genomics foram identificados 117.450 SNP em 72 indivíduos de *C. canephora*. Após análises de qualidade, 33.485 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. canephora* e com base em nessas análises foi possível selecionar possíveis cruzamentos promissores dentro e entre os grupos varietais e Híbridos geneticamente mais distantes. Com base nas análises utilizando o método REML/BLUP foi possível selecionar genótipos Conilon e Robusta promissores por serem geneticamente divergentes pela análise de agrupamento e por proporcionarem ganhos elevados com a seleção. Estes podem ser inter cruzados para obter cultivares dentro de cada grupo varietal e/ou como genitores em cruzamentos interpopulacionais. Além disso, foi possível selecionar famílias híbridas que se destacaram nas análises. Estas podem ser testadas em diferentes regiões e as promissoras utilizadas como variedade propagada por semente. As famílias híbridas selecionadas podem também ser inter cruzadas (cruzamento intrapopulacional) para avanço da seleção recorrente. Por fim, a GWS mostrou-se ferramenta útil para o melhoramento de *C. canephora*, por prever com alta 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 por unidade de tempo.

ABSTRACT

ALKIMIM, Emilly Ruas, D.Sc., Universidade Federal de Viçosa, July, 2017. **Genetic diversity, selection gains and genome wide selection in the *Coffea canephora* species.** Advisor: Eveline Teixeira Caixeta. Co-advisors: Marcos Deon Vilela de Resende and Felipe Lopes da Silva.

The presence of genetic variability associated with more efficient selection strategies is essential for success in breeding programs. In this sense, the use of molecular markers in diversity studies, as well as selection based on phenotypic data through the use of more refined genetic-statistical procedures have been shown to be useful tools in breeding programs. In addition, with the advancement of NGS (Next Generation Sequencing) platforms, a new selection method, which aims to use phenotype and molecular data, called the Genome Wide Selection was proposed. The objective of this work was to identify SNP markers (Single Nucleotide Polymorphism) and to validate them in studies of genetic diversity; to estimate genetic parameters and selection gains using the mixed model methodology (REML/BLUP); to apply the principle of GWS (Genome Wide Selection) and to evaluate its efficiency in *C. canephora* population. The study population was initially composed of 72 genotypes in which the SNP markers were identified and validated. Phenotypic selection, using the REML/BLUP procedure, was performed on 192 coffee genotypes. Finally, the population evaluated through GWS consisted of 165 genotypes. Using the RAPiD Genomics company sequencing platform, 117,450 SNPs were identified in 72 *C. canephora* individuals. After quality analyzes, 33,485 SNPs were selected and validated in analyzes of diversity and genetic structure of populations. The markers were efficient in evaluating the genetic diversity and structure of *C. canephora* and based on these analyzes, it was possible to select possible promising crosses within and among the genetically distant varietal and hybrid groups. Based on the analyzes using the REML/BLUP method, it was possible to select promising Conilon and Robusta genotypes because they are genetically divergent by cluster analysis and because they provide high gains with selection. These can be cross-linked to obtain cultivars within each varietal group and/or as parents at interpopulation crossings. In addition, it was possible to select hybrid families that stood out in the analyzes. These can be tested in different regions and the promising ones used as seed propagated variety. Selected hybrid families can also be cross-linked (intrapopulation cross) to advance recurrent selection. Finally, the GWS proved to be a useful tool for the improvement of *C. canephora*, by accurately predicting the genomic genetic values of the individuals and allowing the reduction in the time needed to complete the selection cycle, providing significant gains in selective efficiency by unit of time.

Introdução geral

O café é a segunda commodity no comércio mundial. Pertencente ao gênero *Coffea*, as duas espécies de maior importância comercial são *Coffea arabica* e *Coffea canephora* (Vega et al. 2003; Denoed et al. 2014). O café arábica é responsável pelo aroma e sabor adocicado, proporcionando bebida de melhor qualidade. O café canephora apresenta maiores quantidades de cafeína, sólidos solúveis e oferece corpo à bebida (Babova et al. 2016).

A espécie *C. canephora* é diploide, alógama e apresenta sistema de autoincompatibilidade gametofítica. Nos programas de melhoramento genético do cafeeiro, essa espécie se destaca pela maior variabilidade genética em relação à *C. arabica*, rusticidade, elevado potencial produtivo, e resistência ao agente causal da ferrugem alaranjada do café, principal doença dessa cultura (Maia et al. 2017).

A diversidade dessa espécie foi descrita por Berthaud (1986), que a dividiu em dois grupos de acordo com seus centros de diversidade. O grupo Guineano compreende os genótipos do oeste africano e o grupo Congolês, os genótipos da região central da África. O grupo Congolês ao longo do processo de domesticação e melhoramento foi difundido para outros locais e estruturado em cinco subgrupos denominados SG1, SG2, B, C, e mais recentemente, um novo subgrupo (UW) composto por acessos selvagens de Uganda (Musoli et al. 2009), justificando a maior variabilidade genética presente nessa espécie.

O subgrupo SG1 reúne os genótipos que ocorrem do Benin ao Gabão, denominados café Conilon, que se destacam por ser subgrupo de melhor adaptação no Brasil e está presente nas principais variedades desenvolvidas no país. Os subgrupos SG2 (oriundos da região da República Democrática do Congo), B (República Centro-Africana) e C (Camarões), compreendem os genótipos conhecidos como café Robusta. Tais cafeeiros compreendem plantas mais altas, vigorosas, de folhas e frutos maiores, resistentes à ferrugem do cafeeiro e mais sensíveis à seca (MARRACCINI et al. 2012).

Em programas de melhoramento genético a presença de variabilidade genética é condição básica e necessária para se obter ganhos genéticos com seleção (Babova et al. 2016). Associado a isso, estratégias eficientes de melhoramento e de seleção de plantas que permitam identificar genótipos superiores são imprescindíveis. Nesse sentido, o uso de marcadores moleculares tem se mostrado ferramenta útil no estudo e caracterização da diversidade genética. Além disso, a seleção baseada em dados fenotípicos também tem sido utilizada com sucesso, por meio do uso de procedimentos genético-estatísticos mais refinados e precisos na estimação de valores genéticos como a metodologia de modelos mistos do tipo REML/BLUP (*Residual or Restricted Maximum Likelihood/Best Linear Unbiased Prediction*).

Com o aprimoramento das técnicas de biologia molecular e das plataformas de NGS (*Next Generation Sequencing*), foi proposto um novo método de seleção, que associa os dados fenotípicos aos dados genotípicos, denominado Seleção Genômica Ampla (*Genome Wide Selection - GWS*) (Meuwissen et al. 2001). A GWS, por meio de marcadores moleculares, enfatiza a predição simultânea dos efeitos genéticos dos *loci* que governam as características de interesse (Almeida et al. 2016). Entretanto, apesar da expressiva importância da espécie *C. canephora* no cenário mundial do agronegócio do café, nenhum estudo com aplicação da GWS para essa espécie foi publicado até o momento.

Referências

- Almeida ÍF de, Cruz CD, Resende MDV de (2016) Validação e correção de fenótipos na seleção genômica ampla. *Pesqui Agropecuária Bras* 51:1973–1982. doi: 10.1590/s0100-204x2016001200008
- Babova O, Occhipinti A, Maffei ME (2016) Chemical partitioning and antioxidant capacity of green coffee (*Coffea arabica* and *Coffea canephora*) of different geographical origin. *Phytochemistry* 123:33–39. doi: 10.1016/j.phytochem.2016.01.016
- Berthaud J (1986) Les ressources génétique pour l'amélioration des caféiers africains diploides. Evaluation de la richesse génétique des populations sylvestres et de ses mécanismes organisateurs. Conséquences pour l'application. Université de Paris
- Denoeud F, Carretero-Paulet L, Dereeper A, et al (2014) The coffee genome provides insight into the convergent evolution of caffeine biosynthesis. *Science* (80-) 345:1181–1184. doi: 10.1126/science.1255274
- Maia T, Badel JL, Fernandes MB, et al (2017) Variation in Aggressiveness Components in the *Hemileia vastatrix* Population in Brazil. *J Phytopathol* 165:174–188. doi: 10.1111/jph.12548
- Marraccini P, Vinecky F, Alves GSC, et al (2012) Differentially expressed genes and proteins upon drought acclimation in tolerant and sensitive genotypes of *Coffea canephora*. *J Exp Bot* 63:4191–4212. doi: 10.1093/jxb/ers103
- Meuwissen T, Hayes B, Goddard M (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157:1819–1829. doi: 11290733
- Musoli P, Cubry P, Aluka P, et al (2009) Genetic differentiation of wild and cultivated populations: diversity of *Coffea canephora* Pierre in Uganda. *Genome* 52:634–646. doi: 10.1139/G09-037
- Vega FE, Rosenquist E, Collins W (2003) Global project needed to tackle coffee crisis. *Nature* 425:343. doi: 10.1038/425343a

CAPÍTULO 1

**High-throughput targeted genotyping and validation of SNP markers in *Coffea canephora*
using next generation sequencing**

ABSTRACT

The development and use of SNP markers for coffee have provided new perspectives for the evaluation of genetic diversity and population structure via different statistical approaches. In this study, 72 *Coffea canephora* genotypes were analyzed in order to identify the SNP markers and apply them in genetic studies and for the selection of parents/hybrids in genetic breeding. As many as 117,450 SNPs were identified using the RAPiD Genomics platform. After quality analyses, 33,485 SNPs were validated for analyses of genetic diversity and population structure. Genotypes were separated according to their varietal groups, and Hybrids were differentiated using clustering and Bayesian approach. Coffee accessions mistakenly identified in the germplasm and in the breeding program were detected. The Conilon varietal group presented the lowest genetic dissimilarity values, and introduction of new accessions in the germplasm was suggested. The highest values of genetic distances were observed among genotypes of the heterotic groups, Conilon and Robusta. The markers were efficient in evaluating the genetic diversity and population structure of *C. canephora*, and promising crosses were selected within and between the varietal groups. Hybrids with greater distances, which were important for *C. canephora* breeding programs, were selected.

Keywords: Single nucleotide polymorphism; genetic variability; population structure; Conilon; Robusta; Hybrid.

Introduction

The use of molecular markers for the detection of polymorphism at DNA level is one of the most significant developments in molecular biology techniques. Molecular markers are more advantageous than phenotypic markers for being uninfluenced by the environment. In addition, they can also be used at the plant development stage (Gartner et al. 2013).

Among the several types of molecular markers, SNPs have stood out for being the most abundant type of polymorphism in genomes, for being codominant and biallelic, and for allowing the automation of the technique of obtaining these markers (Resende et al. 2008; Liao and Lee 2010). In recent years, these markers have been frequently used in studies on humans (Gabriel et al. 2002; Ojopi et al. 2004; Pinto et al. 2008; Gnirke et al. 2009) and in some animal and plant species (Abatepaulo et al. 2008; Caetano 2009; Yang et al. 2013) (Zhang et al. 2011; Neves et al. 2013; Neves et al. 2014; Zhang et al. 2015; Resende et al. 2016). However, these markers have not yet been identified in several plant species of economic importance, such as the coffee species *Coffea canephora*, as they are relatively recent.

With the development of new Next-Generation Sequencing (NGS) technologies, the discovery of SNP became easier and faster and the costs of datapoint were reduced (Carvalho and Silva 2010).

A methodology for obtaining SNPs using NGS has been developed in humans (Gnirke et al. 2009) and adapted in plants (Neves et al. 2013; Neves et al. 2014) by the company RAPiD Genomics located in Florida, USA. This technology makes use of a method of genotyping-by-sequencing of specific regions of the genome. Using this strategy, all regions of interest and/or the entire genome can be analyzed using specific probes used in the analysis.

This new technique provides SNP markers for the species *C. canephora*. Once obtained, the SNPs can be used in several relevant studies of the species, such as in the analyses of genetic diversity and population structure. These analyses enable the validation of SNPs to evaluate the discriminatory capacity in genetically related populations, besides the genetic variability and population structure of the species studied.

The presence of genetic variability is a basic and necessary condition for success in a genetic breeding program. The genetic variability observed in *C. canephora* species is mostly due to the large geographic distribution (especially of the genus *Coffea*), the domestication process; the natural allogamy observed in the species (Davis et al. 2006; Babova et al. 2016).

The diversity of *C. canephora* was first described by Berthaud (1986). The author identified two distinct genetic groups based on their respective centers of diversity: the Guinean

group, which consists of West African genotypes (Guinea and Ivory Coast), and the Congolese group, which consists of genotypes from Central Africa. Other studies have also indicated that the division of *C. canephora* into two groups is strongly related to its geographic isolation and historical events of glaciations that occurred 18,000 years ago (Gomez et al. 2009; Cubry et al. 2013).

The Congolese group, unlike the Guinean group, was disseminated and progressively improved throughout the processes of domestication and breeding of *C. canephora*. The Congolese group was separated into five subgroups, namely SG1, SG2, B, C, and UW (recently discovered and composed of wild accessions from Uganda) (Musoli et al. 2009).

The subgroup SG1 is composed of the genotypes that occur from Benin to Gabon, also known as Conilon coffee. SG1 stands out as the subgroup that is best adapted to Brazil and is present in the main varieties developed in the country. The subgroups SG2 (from the Democratic Republic of Congo), B (from Central African Republic), and C (from Cameroon) consist of the genotypes known as Robusta coffee. These coffee plants are tall, vigorous, possess large leaves and fruits, are resistant to coffee rust, and are more susceptible to drought (Marraccini et al. 2012).

Conilon and Robusta are considered two heterotic groups with distinct and complementary traits within the *C. canephora* species. Thus, the expressive genetic variability observed between and within *C. canephora* population makes this species extremely important in breeding programs for obtaining new varieties that benefit from heterosis (Lashermes et al. 2000; Leroy et al. 2005).

Rapid advances in molecular technologies have opened new perspectives for the evaluation of genetic diversity, involving different statistical approaches (Ferrão et al. 2015). Among these approaches, a genetic dissimilarity matrix (or similarity) can be constructed through the analysis of all possible pairwise genotypic combinations, which can subsequently be used to establish patterns of similarities and dissimilarities (Kosman and Leonard 2005). A Bayesian approach implemented in the STRUCTURE software (Pritchard et al. 2000) has also been used to study the genetic diversity and population structure. In this software, genotypic data are used for the probabilistic classification of genotypes considering K populations (in which K may be unknown).

In view of the above, the objective of the study was to identify SNP molecular markers for *C. canephora* species, validate them for genetic studies, and use them for selection of parents and hybrids for breeding.

Material and methods

Genetic material

An interpopulational partial diallel was carried out for the formation of base populations to initiate the recurrent selection program of *C. canephora* in the breeding program of the Agricultural Research Company of Minas Gerais (Epamig), in association with the Federal University of Viçosa (UFV) and the Brazilian Company of Agricultural Research - Coffee (Embrapa Café). Five genotypes of Conilon (male parent) and five genotypes of Robusta (female parent) were used in the diallel. Besides these parents, five Conilon genotypes, four Robusta genotypes (Table 1), as well as 20 Hybrids (Conilon × Robusta) (Table 2), which were obtained from the diallel, were also included in the study population. Twenty-one genotypes of the Conilon varietal group and 22 of the Robusta varietal group (Table 3) of the Epamig/UFV/Embrapa breeding program were selected, totaling 72 genotypes.

The population was selected to represent a broad genetic variability of the species *C. canephora*. To this end, divergent materials with agronomic traits of interest were selected.

Genomic DNA extraction and SNP identification

Young and fully expanded leaves from the 72 genotypes were collected, and the genomic DNA was extracted using the methodology described by (Diniz et al. 2005). The DNA concentration was verified in NanoDrop 2000, and its quality was evaluated in 1% agarose gel. The DNA concentration of the samples was standardized and sent to RAPiD Genomics for the construction of probes, sequencing, and identification of SNP molecular markers.

For the identification and genotyping of the SNP markers, 40,000 probes were used (Resende et al. 2016). The probes were constructed from reference sequences. The database of the Brazilian Project of Coffee Genome was used in this study. It contains more than 200,000 Expressed Sequence Tags (ESTs), corresponding to approximately 33,000 transcribed genes known as Unigenes (Vieira et al. 2006). The database of the *C. canephora* genome was also used, containing a total of 25,574 genes (Denoeud et al. 2014). Using these reference sequences, specific probes were obtained to cover the entire genome, considering both transcribed and non-transcribed regions.

The 72 genotypes were sequenced with these probes using the Illumina platform, and the SNP markers were identified using bioinformatics tools (Resende et al. 2016).

Quality analysis of SNPs

The VCFtools software (Danecek et al. 2011) was used for quality analysis of the SNPs (Resende et al. 2016). The following quality parameters were used: MinDP - minimum coverage per SNP and per individual, in which SNPs with values lower than that determined in the filter were considered as missing data; DPrange - mean coverage range of the SNPs in the population; if the mean coverage of a SNP sequenced in a population is lower or higher than that determined by the band, this SNP is removed; Miss - missing data; if the population presents a value higher than that defined as missing (Miss) data, these SNPs are removed; MAF - minor allele frequency; if the frequency of the allele is lower than that of the defined value, the SNP is removed; and MinQ - minimum quality of the SNPs based on the number of read and the quality of the alignment of these reads.

To obtain the dataset of the filtering, the parameters MinDP5, DPrange20-350, Miss0.2, MAF0.03, and MinQ20 were considered.

Analysis of genetic diversity and population structure

The SNP data were coded as 0, 1, and 2 for being biallelic and codominant markers. Thus, considering the genotypes of the individuals as A^1A^1 , A^1A^2 , or A^2A^2 in each locus, these genotypes were coded as 0, 1, or 2, respectively.

The genetic dissimilarity matrix was obtained by the arithmetic complement of the unweighted index using the GENES software (Cruz 2013). Genetic distance can be estimated by the following equation:

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

where:

$D_{ii'}$ = genetic distance between pairs of accessions i e i' ;

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

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

a_j : total number of loco alleles j ;

A : total number of alleles studied;

c_j : number of common alleles between pairs of accessions i e i' .

The cluster analysis for the construction of dendrograms was performed using the Unweighted Pair Group Method with Arithmetic Mean (UPGMA) and Neighbor Joining (NJ), using the statistical software MEGA 7.0 (Kumar et al. 2016).

Another cluster analysis was performed using the STRUCTURE software (Pritchard et al. 2000), which considered the K values ranging from 1 to 6, with 10 replicates. Each run was carried out with a 25,000 burn-in period, followed by 75,000 MCMC (Markov Chain Monte Carlo). The best number of genetic groups was estimated by calculating the ΔK , being this obtained by the highest value of ΔK , as described by Evanno et al. (2005), by using the software STRUCTURE HARVESTER (Earl and VonHoldt 2012).

Principal Components Analysis (PCA) is performed using the `pcaMethods` R package (Stacklies et al. 2007) implemented in ClustVis (Metsalu and Vilo 2015).

Results

As many as 117,450 SNPs were identified through the sequencing of 72 *C. canephora* genotypes (26 Conilons, 26 Robustas, and 20 Hybrids) by using 40,000 probes distributed in the genome. After the quality analyses of the SNP, 33,485 SNP markers were obtained with the filter.

The genetic distance matrix was generated using the molecular data, and two large groups (I and II) were formed with a dendrogram, by using the UPGMA technique. The first group consisted of the accessions of the varietal group Robusta, and the second group consisted of the accessions of the varietal group Conilon and the Hybrids (Conilon \times Robusta). The second group can be divided into two subgroups, II.a and II.b, consisting of the accessions of Conilon and Hybrids, respectively (Figure 1).

The Hybrid H0915-1 was allocated in the group of Robusta accessions. The genotypes Conilon UFV 3628-2 and UFV 3628-45 and the genotype Robusta UFV 3374-29 were allocated in the subgroup of Hybrids (Figure 1).

The results obtained using the NJ algorithm for cluster analysis were equivalent to those obtained using the UPGMA methodology. The two dendrograms allocated the genotypes in the Conilon, Robusta, and Hybrid groups, respectively. The Hybrid H0915-1 was also allocated in the group of Robusta genotypes. The genotypes Conilon UFV 3628-2 and UFV 3628-45 and the genotype Robusta UFV 3374-29 were allocated closer to the subgroup of the Hybrids (Figure 2).

The Bayesian approach performed using the STRUCTURE software (Pritchard et al. 2000) confirmed the results found using the clustering methods. According to the methodology proposed by Evanno et al. (2005), the highest value of ΔK was obtained for $K = 2$ (Figure 3), showing that

the genotypes can be separated into two groups according to the Conilon or Robusta varietal group (Figure 4a).

The Hybrid H0915-1 presents approximately 80% genetic similarity with Robusta genotypes. The genotypes Conilon UFV 3628-2 and UFV 3628-45 showed approximately 50% genetic similarity with the Conilon group and 50% genetic similarity with the Robusta group. The genotype Robusta UFV 3374-29 showed approximately 80% genetic similarity with the Robusta group and 20% genetic similarity with the Conilon group (Figure 4a).

Considering the division of the population into three groups ($K = 3$) (Figure 4b), the Robusta genotypes were allocated to group I (red bars). In group II (blue bars) the Hybrids as well as the genotypes Conilon UFV 3628-2 and UFV 3628-45 and the genotypes Robusta UFV 3368-58 and UFV 3374-29 were allocated. In group III (green bars) the Conilon genotypes were allocated. The Hybrid H0915-1 was allocated to group II, but presented about 38% of genetics mixture with group I. The genotype Robusta UFV 3368-58 was allocated to group II, but presented about 42% of genetics mixture with group I. The genotype Robusta UFV 3374-29 was allocated to group II, but presented about 31% of genetics mixture with group I and 5% of genetics mixture with group III.

Values of genetic dissimilarities estimated between and within the varietal groups (Conilon and Robusta) and the Hybrids are presented in Tables 4 and 5. Table 4 shows the lowest genetic dissimilarities observed and the mean distance between and within the groups. Table 5 shows the five major genetic dissimilarities obtained between and within the varietal groups and the Hybrids.

The lowest estimates of genetic distances were obtained among the genotypes of the Conilon varietal group, and the lowest distance (0.0071) was observed between the accessions Conilon UFV 3628-45 and UFV 3628-2 (Table 4). These genotypes were grouped in the Hybrids subgroup, according to the analyses of the dendrograms.

The highest values of genetic distances were observed between the genotypes of the Conilon and Robusta varietal groups. The highest distance (0.3923) was obtained between the genotypes Conilon UFV 3629-11 and Robusta UFV 3371-19 (Table 5). The lowest genetic distance was obtained between the genotypes Conilon UFV 3628-45 and Robusta UFV 3374-29. According to the analyses of the dendrograms, the two genotypes were allocated in the Hybrid subgroup. The overall mean genetic distance observed was 0.2395.

The Hybrids H0915-1, H0919-2, and H0921-1 presented the highest values of genetic dissimilarity compared to other Hybrids and the genotypes of the Conilon and Robusta groups. Among these three Hybrids, the greatest genetic distance was observed between H0915-1 and

H0921-1 (0.2284) (Table 5). According to the analyses of the dendrograms, the Hybrid H0915-1 was allocated in the Robusta group (Figure 1 and 2).

Thus, disregarding the genotypes Conilon UFV 3628-2 and UFV 3628-45, the genotype Robusta UFV 3374-29, as well as the Hybrid H0915-1 mistakenly allocated in their respective groups, the five largest distances between and within the varietal groups and Hybrids are described in Table 6.

Based on the data in Table 6, some genotypes of the Conilon and Robusta varietal groups with greater genetic dissimilarity can be selected for the formation of base populations. Hybrids of interest for the *C. canephora* breeding program can also be selected, because by recommending a set of Hybrids with greater genetic distances, the problems with genetic autoincompatibility present in the species *C. canephora* are avoided. Therefore, Hybrids H0911-2, H0919-2, H0922-1, H0912-1, H0921-1, and H0918-1 are found to be promising for to be recommended as a set of Hybrid varieties because they present the highest genetic distance values.

A PCA was performed using SNP markers and two principal components were observed (Figure 5). PC1 and PC2 contributed variations of 68.4% and 19.0%, respectively. Cumulatively these two PCs contributed 87.4% of total variation observed. The analysis of PCA showed the formation of three groups. The genotypes Conilon UFV 3628-45, UFV 3628-2 and UFV 3629-11, the genotype Robusta UFV 3368-58 and the Hybrid H0915-1 were not allocated to any group and the genotype Robusta UFV 3374-29 was allocated to the Hybrids group. The genotypes Conilon UFV 3628-45 and UFV 3628-2, although not located in any group, were closer to the Hybrid group and Hybrid H0915-1 was close to the Robusta group. The genotype Robusta UFV 3368-58, even though it has not been allocated in any group, this is located between to the Robusta and Hybrid groups corroborate with the results obtained using the software STRUCTURE for K= 3 (Figure 4b). These results corroborate with the results obtained in the previous analyzes, except for genotype Conilon UFV 3629-11. However, this genotype, even though it has not been allocated in any group, by the PCA analysis, this is located close to their respective group. The results also show that the Robusta genotypes have greater spatial dispersion (genetic variability) in relation to the genotypes of Conilon.

Discussion

To ensure that the entire genome was covered and that probes were obtained in transcribed and non-transcribed regions, 40,000 specific probes were constructed by means of reference sequences. The probes were constructed in order to obtain a gene probe annotated from the *C.*

canephora genome. The probes for non-gene regions were selected so that two probes were present in a 10 kb interval. The other probes were from the Brazilian Coffee Genome Project (databases of ESTs of the species *C. arabica* and *C. canephora*) (Resende et al. 2016).

Due to the large number of data obtained from genotyping with thousands of molecular markers, even a low error rate can be detrimental in subsequent analyses. Thus, some parameters for quality control and data filtering must be used in order to remove these errors. These evaluations allow the identification of below-standard markers and samples, which must be removed prior to analysis in order to reduce false positive error (type I error) and false negative error (type II error) (Anderson et al. 2010).

The RAPiD Genomics platform presented a large potential for SNP identification and genotyping, since a significant number of these markers was identified in *C. canephora* compared to other studies (Zhou et al. 2016; Garavito et al. 2016). These markers were able to separate the genotypes of the studied population into distinct groups. Thus, the efficiency of these SNP markers has been proven in the analyses of genetic diversity and population structure.

A technique based on the reduction of complexity by using the restriction enzymes that aim at gene-rich regions and by using the NGS technology, known as DArTseq, was applied to test the performance of the markers derived from this method in coffee plants. By studying the genetic diversity of the species *C. canephora* cultivated in Vietnam and Mexico, 4,021 SNPs were identified (Garavito et al. 2016). These SNP markers constitute a valuable tool for breeders. In addition, they increase the knowledge on the genetic diversity of *C. canephora* and contribute to the understanding of the genetic background of important varieties from major coffee producers (Garavito et al. 2016).

In another study, a set of SNP markers was developed and used to assist in the identification of coffee germplasm, demonstrating the usefulness of EST sequences as an approach for the rapid development of a high-throughput targeted genotyping tool (Zhou et al. 2016). By using ESTs of the *C. arabica*, *C. canephora*, and *C. racemosa* obtained from public databases, 7,538 SNP markers were developed, of which 180 were selected for validation using 25 accessions of *C. arabica* and *C. canephora* from Puerto Rico. Based on the validation result, a panel of 55 SNP markers that are polymorphic between both coffee species was constructed (Zhou et al. 2016).

In the UPGMA clustering, the SNPs were able to classify the genotypes into different groups according to the varietal group (Conilon and Robusta). It also discriminated the Hybrid genotypes obtained from the cross between Conilon and Robusta genotypes. The results were confirmed by the analysis using the NJ algorithm, the Bayesian approach performed using the

STRUCTURE software (Pritchard et al. 2000) and the PCA analysis. These three analyses were also efficient in separating the Conilon and Robusta varietal groups. However, as in the previous UPGMA cluster analysis, the Hybrid genotypes were also discriminated.

Robusta and Conilon heterotic groups present distinct and striking phenotypic traits. Nevertheless, the classification of these materials is no easy task. This is due to the strictly allogamous reproduction form, high phenotypic amplitude, and heterozygosity, in addition to possible natural crosses between the two varietal groups. Thus, studies on the genetic diversity using molecular markers are crucial in breeding programs for the conservation of genetic sources and for the correct evaluation of genotypes (Ferrão et al. 2013; Ren et al. 2013; Ferrão et al. 2015).

Three accessions, Conilon UFV 3628-2, Conilon UFV 3628-45, and Robusta UFV 3374-29, were not allocated in their respective varietal groups but in the subgroup of the Hybrids. This result suggests that these accessions are probably derivatives of natural crossings and were mistakenly classified.

The two Conilon genotypes (UFV 3628-2 and UFV 3628-45), as well as the other Conilon genotypes are derived from half-sib seeds. Since Conilon is allogamous and has an autoincompatibility system, the two seeds that originated these two genotypes may have been originated from the pollination with a Robusta parent, instead of Conilon, forming natural Hybrids in the location where they were collected.

The genotype Conilon UFV 3628-2 was chosen as the parent to obtain Hybrids in the interpopulational partial diallel due to its prominence in yield, resistance to rust, and earliness. The genotype Conilon UFV 3628-45 also stood out in the evaluation of these traits. Resistance to rust is an evidence that suggests that these accessions may be natural Hybrids, since Conilon genotypes are usually susceptible to rust. Resistance to this disease is usually obtained from Robusta (Cubry et al. 2008; Musoli et al. 2009). In addition, relatively high yield was observed in these accessions compared to the other Conilon genotypes. In the mean of five years of evaluations (2012 to 2016) (data not shown), the genotypes Conilon UFV 3628-2 and Conilon UFV 3628-45 were in the fifth and eighth position, respectively, in yield data compared to the other Conilon genotypes evaluated in the breeding program. Yield increase may also be the result of Hybrid vigor, corroborating the molecular data that indicate these natural Hybrid genotypes.

The genotype Robusta UFV 3368-58 is probably derivative of natural crossings and also was mistakenly classified. Using the software STRUCTURE for $K=3$ this genotype was allocated to group II, in the analysis using the NJ algorithm it is also located close to the Hybrid genotypes

and in the PCA analysis, even though it has not been allocated in any group, this genotype is located between to the Robusta and Hybrid groups.

The accession Robusta UFV 3374-29, also indicated as a probable natural Hybrid by the molecular data, stood out in relation to the mean yield of the five years of evaluations, although data were slightly inferior compared to the accessions Conilon UFV 3628-2 and Conilon UFV 3628 -45. This accession stands out in relation to its earliness, which was probably obtained from its possible parent Conilon. In the analysis using the STRUCTURE software for $K= 2$, approximately 78% similarity is observed with Robusta and approximately 22% similarity is observed with Conilon. This contribution percentage of the Conilon parent appears to be sufficient to guarantee the earliness of the genotype Robusta UFV 3374-29. In the analysis using the software STRUCTURE for $K= 3$ the genotype Robusta UFV 3374-29 was allocated to group II, proving that this genotype is a natural Hybrids.

The Hybrid H0915-1 was allocated in the Robusta group in both cluster analyses. According to molecular data, one of the parents of Hybrid H0915-1 is the genotype Conilon 3628-2, which is a natural Hybrid. Thus, this Hybrid was formed by the cross [(Robusta \times Conilon) \times Robusta]. This explains why the Hybrid H0915-1 was genetically closer to Robusta genotypes in the cluster analyses. This result was confirmed using the STRUCTURE software for $K=2$, which shows approximately 84% similarity with Robusta and approximately 16% similarity with Conilon. Using the STRUCTURE software for $K=3$ the Hybrid H0915-1 was allocated to group II, but presented about 38% of genetics mixture with group I, which comprises Robusta genotypes.

The genetic distance matrix showed that the Conilon varietal group presented the lowest values of genetic dissimilarity. The presence of genetic variability within the base population is responsible for the success of the breeding programs (Grandillo 2014). Thus, in order to obtain the base population of genotypes of the Conilon varietal group, genotypes with the highest values of genetic distance observed in the distance matrix must be selected. Therefore, in this germplasm, the five crosses presented in Table 6 should be prioritized, if the objective is to obtain crosses only between accessions of the Conilon group. The possibility of increasing the variability of this group in the Germplasm Bank, by introducing new accessions, should also be considered.

For the formation of the base population of the Robusta varietal group, the accessions presented in Table 6 must also be prioritized, since they present the highest values of genetic dissimilarity and are important sources of variability for breeding programs.

The highest values of genetic distance between genotypes belonging to the Conilon and Robusta heterotic groups were already expected. Therefore, the highest distance value (0.3923)

observed between the genotype Conilon UFV 3629-11 and the genotype Robusta UFV 3371-19 is explained for these genotypes that present distinct and complementary characteristics.

In breeding programs, such as Epamig/UFV/Embrapa, which aim at identifying heterotic groups and contrasting parents to explore Hybrid vigor, the correct evaluation of diversity is a key factor for the choice of parents. The expectation that divergent or contrasting parents generate good Hybrids is due to the fact that if two parents are genetically close, there is a tendency for them to share many common genes or alleles. Thus, while crossing them, little complementarity and low Hybrid vigor will occur due to the low level of allelic heterozygosity in the cross and due to the absence of dominance deviations (Ferrão et al. 2013).

One of the main focuses of the Epamig/UFV/Embrapa breeding programs is to recommend to producers a combination of Hybrid coffee seeds that avoid genetic incompatibility problems; are more drought-tolerant than clonal varieties; and are associated with rust resistance, maturation uniformity, and greater fruit size. Thus, by observing the results and disregarding H0915-1, which was allocated in the Robusta varietal group, the H0911-2, H0919-2, H0922-1, H0912-1, H0921-1, and H0918-1 (Table 6) may be remarkable for the breeding program, since they present the highest values of genetic dissimilarity.

The results presented great efficiency for the identification and genotyping of SNP markers, as well as their ability to evaluate the genetic diversity. The most common molecular markers used in diversity studies in the genus *Coffea* are the Random Amplified Polymorphic DNA (Sera et al. 2003), Amplified Fragment Length Polymorphism (Prakash et al. 2005; Brito et al. 2010; Diola et al. 2011), and Simple Sequence Repeat (Prakash et al. 2005; Ferrão et al. 2013).

Conclusions

The methodology of SNP identification and genotyping used in this study presents great potential for the detection and selection of an expressive number of SNP markers.

The markers were efficient in evaluating the genetic diversity and population structure of *C. canephora*. Based on their analysis, promising crosses were selected within and between the varietal groups, according to the *C. canephora* breeding programs.

Genotypes of the Conilon varietal group showed low genetic diversity, which requires the selection of genotypes with the highest values of genetic distance, as well as the amplification of the variability of this group in the Germplasm Bank, by introducing new accessions.

The highest values of genetic distances observed were between genotypes of Conilon and Robusta heterotic groups.

Hybrids with higher values of genetic distances can be selected, which are important for *C. canephora* breeding programs, in order to avoid genetic incompatibility problems.

Table 1. Parents Conilon and Robusta

Genotypes	Description	Genotypes	Description
UFV 513	Genitor Conilon	UFV 3365-144	Genitor Robusta
UFV 3627-31	Genitor Conilon	UFV 3366-139	Genitor Robusta
UFV 3628-2	Genitor Conilon	UFV 3373-36	Genitor Robusta
UFV 3629-11	Genitor Conilon	UFV 3374-28	Genitor Robusta
UFV 3629-25	Genitor Conilon	---	---

Table 2. Description of the crosses to obtain the Hybrids that make up the population

Hybrid genotypes	Crosses
H092-2	UFV 3366-139 x UFV 513
H093-1	UFV 3366-139 x UFV 3629-11
H094-2	UFV 3367-98 x UFV 513
H095-9	UFV 3366-139 x UFV 3627-31
H097-8	UFV 3374-28 x UFV 3627-31
H098-3	UFV 3365-144 x UFV 3629-11
H099-2	UFV 3365-144 x UFV 513
H0910-4	UFV 3367-98 x UFV 3629-11
H0911-2	UFV 3367-98 x UFV 3628-2
H0912-1	UFV 3366-139 x UFV 3628-2
H0913-1	UFV 3365-144 x UFV 3628-2
H0914-2	UFV 3374-28 x UFV 513
H0915-1	UFV 3373-36 x UFV 3628-2
H0916-1	UFV 3367-98 x UFV 3629-25
H0917-1	UFV 3365-144 x UFV 3629-25
H0918-1	UFV 3373-36 x UFV 513
H0919-2	UFV 3373-36 x UFV 3629-11
H0920-1	UFV 3373-36 x UFV 3629-25
H0921-1	UFV 3374-28 x UFV 3629-11
H0922-1	UFV 3373-36 x UFV 3627-31

Table 3. Genotypes of the Conilon and Robusta varietal groups

Conilon Group		Robusta Group	
Genotypes	Genotypes	Genotypes	Genotypes
UFV 3627-20	UFV 3628-37	UFV 514	UFV 3370-47
UFV 3627-24	UFV 3628-45	UFV 3356-71	UFV 3371-19
UFV 3627-27	UFV 3629-4	UFV 3356-74	UFV 3373-43
UFV 3627-29	UFV 3629-7	UFV 3357-93	UFV 3374-29
UFV 3627-30	UFV 3629-10	UFV 3358-88	UFV 3375-65
UFV 3628-1	UFV 3629-17	UFV 3360-169	UFV 3376-8
UFV 3628-3	UFV 3629-27	UFV 3361-148	UFV 3377-12
UFV 3628-5	UFV 3629-29	UFV 3362-118	UFV 3630-2
UFV 3628-16	UFV 3629-30	UFV 3366-134	UFV 3631-1
UFV 3628-24	UFV 3629-34	UFV 3367-101	UFV 3631-6
UFV 3628-29	--	UFV 3368-58	UFV 3631-11

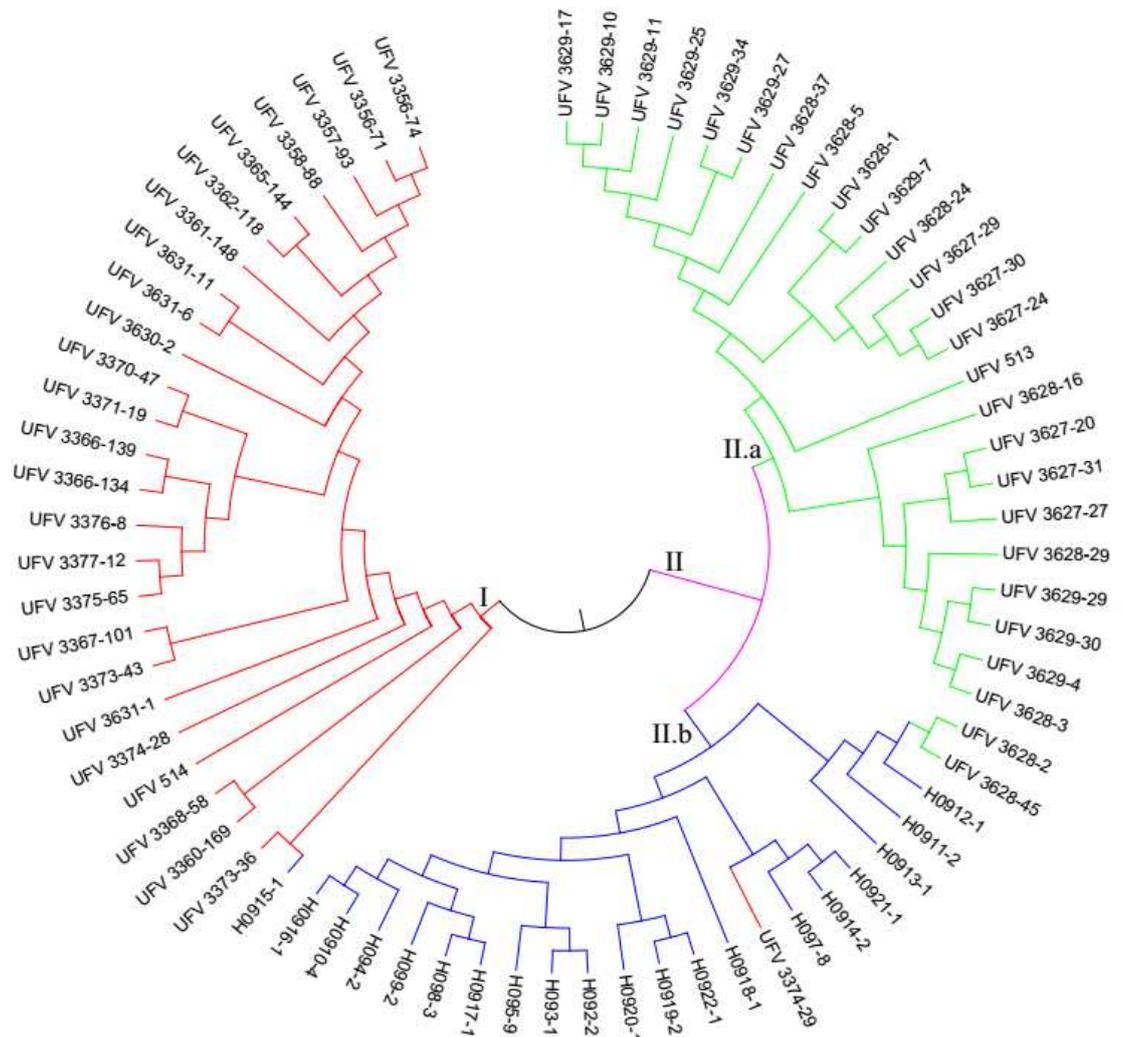


Fig 1 Dendrogram obtained by the UPGMA technique, based on the arithmetic complement of the unweighted index of 72 *Coffea canephora* genotypes (26 Conilons, 26 Robustas, and 20 Hybrids) and 33,485 SNP markers. Green, red and blue lines represent the varietal groups (Conilon and Robusta) and the Hybrids, respectively

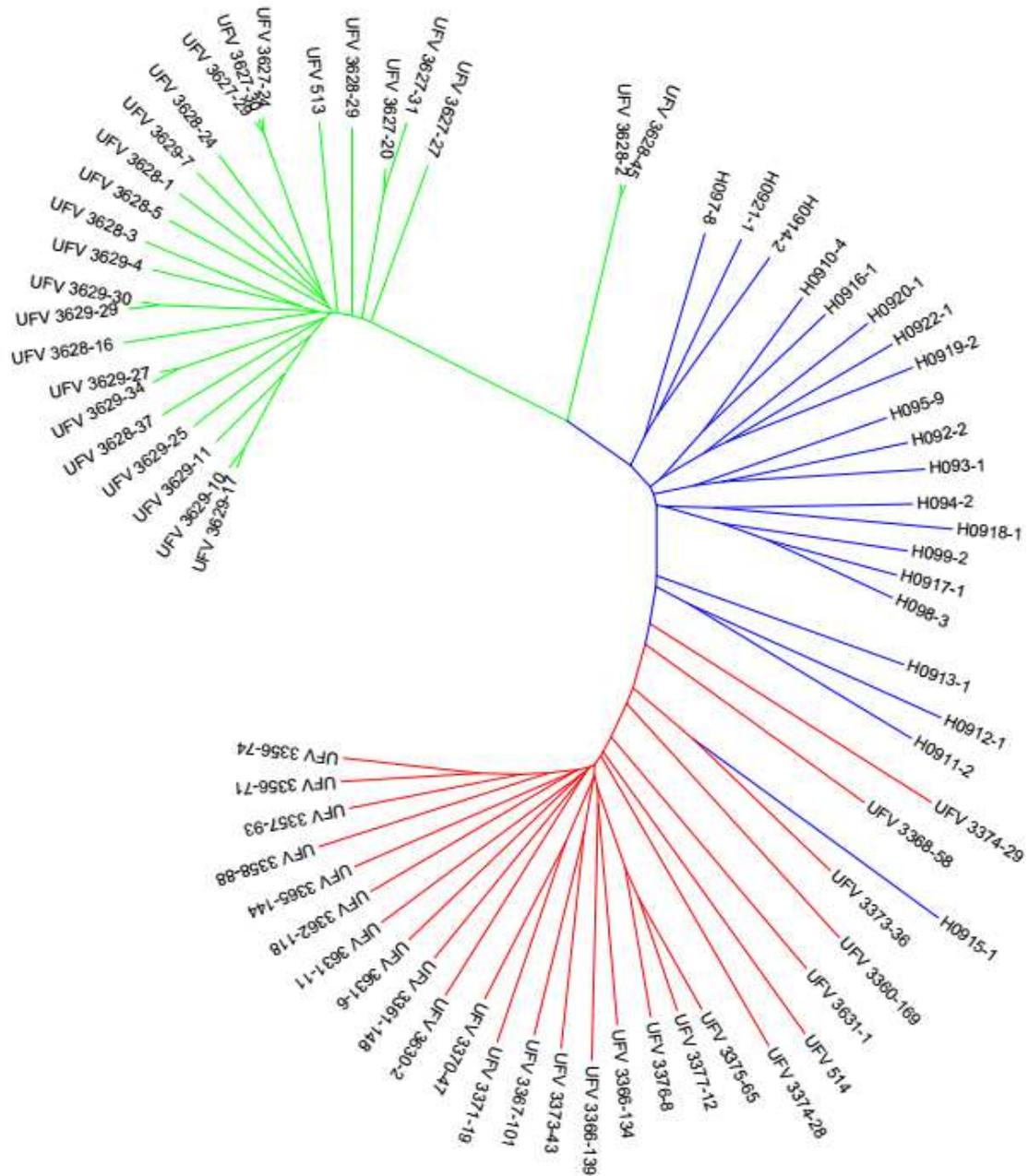


Fig 2 Neighbor Joining Dendrogram (NJ) of the 72 *Coffea canephora* genotypes (26 Conilons, 26 Robustas and 20 Hybrids). Green, red and blue lines represent the varietal groups (Conilon and Robusta) and the Hybrids, respectively

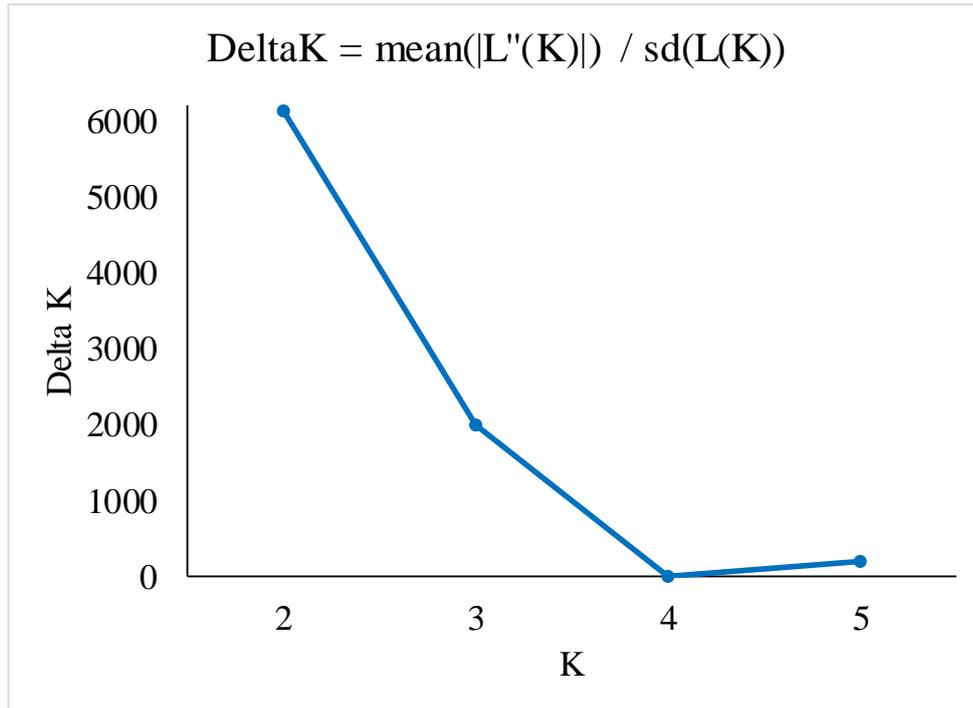


Fig 3 Graphic obtained with the values of ΔK for visualization of the best K (K = 2), according to the methodology proposed by Evanno et al. (2005)

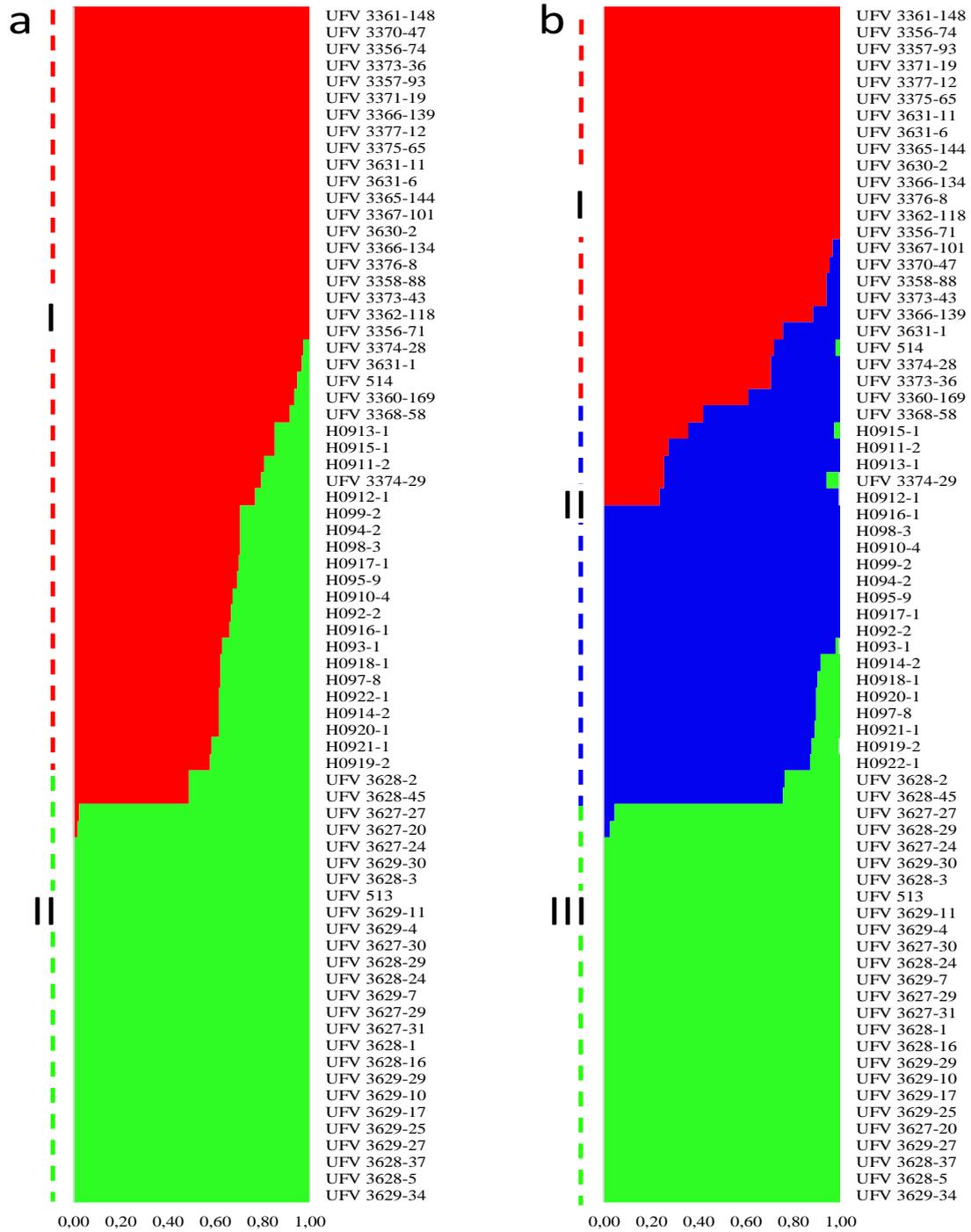


Fig 4 Bar graphic of the STRUCTURE software used to study the diversity of the 72 *Coffea canephora* genotypes (26 Conilons, 26 Robustas, and 20 Hybrids). The 72 genotypes are represented below the graphic, and were divided into two ($k = 2$) and three ($k = 3$) groups, Figures 5a and b respectively, according to the Conilon and Robusta varietal group, and also allowed discriminating Hybrid genotypes resulting from crosses between the two heterotic groups. The red bar is associated to the Robusta genotypes, the green bar to the Conilon genotypes and the blue bar to the Hybrid genotypes

Table 4. Lower values of genetic dissimilarity and mean genetic dissimilarity obtained between and within the varietal groups (Conilon and Robuta) and the Hybrids

Groups	Genotype X	Genotype Y	Dissimilarity	Mean Dissimilarity
Conilon x Conilon	UFV 3628-45	UFV 3628-2	0.0071	0.1226
Robusta x Robusta	UFV 3377-12	UFV 3375-65	0.0756	0.1707
Hybrid x Hybrid	H098-3	H0917-1	0.0827	0.1662
Conilon x Robusta	UFV 3628-45	UFV 3374-29	0.2203	0.3538
Conilon x Hybrid	UFV 3628-2	H0912-1	0.1603	0.2321
Robusta x Hybrid	UFV 3373-36	H0915-1	0.1508	0.2469

Table 5. Greater values of genetic dissimilarity obtained between and within the varietal groups (Conilon and Robusta) and Hybrids

Groups	Genotype X	Genotype Y	Dissimilarity
Robusta x Conilon	UFV 3371-19	UFV 3629-11	0.3923
	UFV 3365-144	UFV 3629-11	0.3911
	UFV 3357-93	UFV 3627-24	0.3907
	UFV 3357-93	UFV 3628-3	0.3905
	UFV 3357-93	UFV 3627-29	0.3900
Hybrid x Conilon	H0915-1	UFV 3627-24	0.3024
	H0915-1	UFV 3627-30	0.3016
	H0915-1	UFV 3629-29	0.2998
	H0915-1	UFV 3628-3	0.2983
	H0915-1	UFV 3628-5	0.2966
Robusta x Hybrid	UFV 3371-19	H0921-1	0.2893
	UFV 3375-65	H0919-2	0.2865
	UFV 3631-6	H0919-2	0.2842
	UFV 3377-12	H0919-2	0.2832
	UFV 3365-144	H0921-1	0.2826
Robusta x Robusta	UFV 3374-29	UFV 3371-19	0.2531
	UFV 3374-29	UFV 3365-144	0.2509
	UFV 3374-29	UFV 3356-71	0.2481
	UFV 3374-29	UFV 3630-2	0.2469
	UFV 3374-29	UFV 3356-74	0.2459
Hybrid x Híbrido	H0915-1	H0921-1	0.2284
	H0915-1	H093-1	0.2272
	H0915-1	H0914-2	0.2247
	H0915-1	H092-2	0.2247
	H0915-1	H098-3	0.2243
Conilon x Conilon	UFV 3628-16	UFV 3628-2	0.2160
	UFV 3628-16	UFV 3628-45	0.2157
	UFV 3627-24	UFV 3628-2	0.2111
	UFV 3629-29	UFV 3628-45	0.2110
	UFV 3627-24	UFV 3628-45	0.2110

Table 6. Greater values of genetic dissimilarity obtained between and within the varietal groups (Conilon and Robusta) and the Hybrids, disregarding the accessions Conilon UFV 3628-2 and Conilon UFV 3628-45, Robusta UFV 3374-29, and the Hybrid H0915-1

Groups	Genotype X	Genotype Y	Dissimilarity
Robusta x Conilon	UFV 3371-19	UFV 3629-11	0.3923
	UFV 3365-144	UFV 3629-11	0.3911
	UFV 3357-93	UFV 3627-24	0.3907
	UFV 3357-93	UFV 3628-3	0.3905
	UFV 3357-93	UFV 3627-29	0.3900
Hybrid x Conilon	H0911-2	UFV 3628-16	0.2900
	H0911-2	UFV 3627-30	0.2881
	H0911-2	UFV 3627-24	0.2879
	H0911-2	UFV 3629-25	0.2861
	H0913-1	UFV 3627-24	0.2844
Robusta x Hybrid	UFV 3371-19	H0921-1	0.2893
	UFV 3375-65	H0919-2	0.2865
	UFV 3631-6	H0919-2	0.2842
	UFV 3377-12	H0919-2	0.2832
	UFV 3365-144	H0921-1	0.2826
Robusta x Robusta	UFV 3371-19	UFV 3368-58	0.2198
	UFV 514	UFV 3360-169	0.2183
	UFV 3365-144	UFV 3368-58	0.2164
	UFV 3368-58	UFV 3630-2	0.2155
	UFV 514	UFV 3368-58	0.2149
Hybrid x Hybrid	H0911-2	H0919-2	0.2202
	H0911-2	H0922-1	0.2196
	H0922-1	H0912-1	0.2183
	H0912-1	H0921-1	0.2172
	H0918-1	H0912-1	0.2151
Conilon x Conilon	UFV 3629-30	UFV 3629-7	0.1351
	UFV 3629-7	UFV 3629-29	0.1312
	UFV 3629-30	UFV 3627-29	0.1304
	UFV 3627-24	UFV 3629-30	0.1304
	UFV 3629-30	UFV 3627-30	0.1299

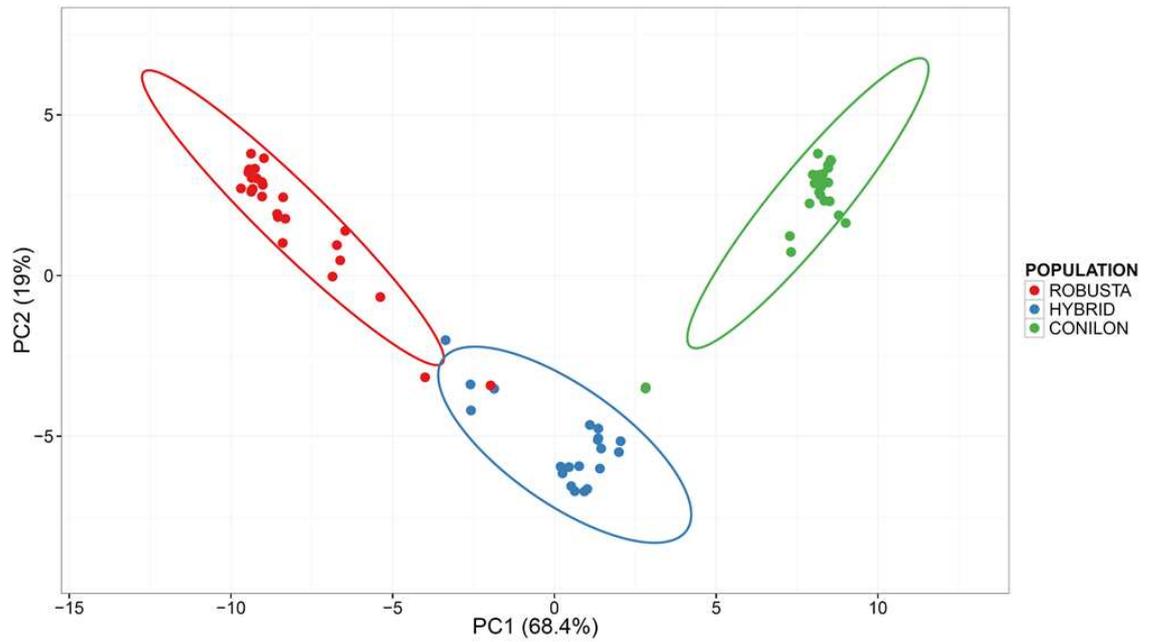


Fig 5 Analysis of Principal Components of the 72 genotypes of *Coffea canephora*, showing the formations of three groups. X and Y axis show principal component 1 and principal component 2 that explain 68.4% and 19% of the total variance, respectively

References

- Abatepaulo ARR, Caetano AR, Mendes Jr CT, et al (2008) Detection of SNPs in bovine immune-response genes that may mediate resistance to the cattle tick *Rhipicephalus (Boophilus) microplus*. *Anim Genet* 39:328–329. doi: 10.1111/j.1365-2052.2008.01713.x
- Anderson CA, Pettersson FH, Clarke GM, et al (2010) Data quality control in genetic case-control association studies. *Nat Protoc* 5:1564–1573. doi: 10.1038/nprot.2010.116
- Babova O, Occhipinti A, Maffei ME (2016) Chemical partitioning and antioxidant capacity of green coffee (*Coffea arabica* and *Coffea canephora*) of different geographical origin. *Phytochemistry* 123:33–39. doi: 10.1016/j.phytochem.2016.01.016
- Berthaud J (1986) Les ressources génétique pour l'amélioration des caféiers africains diploïdes. Evaluation de la richesse génétique des populations sylvestres et de ses mécanismes organisateurs. Conséquences pour l'application. Université de Paris
- Brito GG, Caixeta ET, Gallina AP, et al (2010) Inheritance of coffee leaf rust resistance and identification of AFLP markers linked to the resistance gene. *Euphytica* 173:255–264. doi: 10.1007/s10681-010-0119-x
- Caetano AR (2009) Marcadores SNP: conceitos básicos, aplicações no manejo e no melhoramento animal e perspectivas para o futuro. *Rev Bras Zootec* 38:64–71. doi: 10.1590/S1516-35982009001300008
- Carvalho MC da CG de, Silva DCG Da (2010) Sequenciamento de DNA de nova geração e suas aplicações na genômica de plantas. *Ciência Rural* 40:735–744. doi: 10.1590/S0103-84782010000300040
- Cruz CD (2013) GENES - a software package for analysis in experimental statistics and quantitative genetics. *Acta Sci Agron* 35:271–276. doi: 10.4025/actasciagron.v35i3.21251
- Cubry P, De Bellis F, Pot D, et al (2013) Global analysis of *Coffea canephora* Pierre ex Froehner (Rubiaceae) from the Guineo-Congolese region reveals impacts from climatic refuges and migration effects. *Genet Resour Crop Evol* 60:483–501. doi: 10.1007/s10722-012-9851-5
- 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, Govaerts R, Bridson DM, Stoffelen P (2006) An annotated taxonomic conspectus of the genus *Coffea* (Rubiaceae). *Bot J Linn Soc* 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* (80-) 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 Breed Appl Biotechnol* 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 Genet Genomes* 7:1199–1208. doi: 10.1007/s11295-011-0406-2
- Earl DA, VonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv Genet Resour* 4:359–361. doi: 10.1007/s12686-011-9548-7
- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. *Mol Ecol* 14:2611–2620. doi: 10.1111/j.1365-294X.2005.02553.x
- Ferrão L, Caixeta E, Souza F, et al (2013) Comparative study of different molecular markers for classifying and establishing genetic relationships in *Coffea canephora*. *Plant Syst Evol* 299:225–238. doi: 10.1007/s00606-012-0717-2
- 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. *Mol Breed* 35:31. doi: 10.1007/s11032-015-0247-z
- Gabriel SB, Schaffner SF, Nguyen H, et al (2002) The Structure of Haplotype Blocks in the Human Genome. *Science* (80-) 296:2225–2229. doi: 10.1126/science.1069424
- Garavito A, Montagnon C, Guyot R, Bertrand B (2016) Identification by the DArTseq method of the genetic origin of the *Coffea canephora* cultivated in Vietnam and Mexico. *BMC Plant Biol* 16:242. doi: 10.1186/s12870-016-0933-y
- Gartner GAL, McCouch SR, Moncada MDP (2013) A genetic map of an interspecific diploid pseudo testcross population of coffee. *Euphytica* 192:305–323. doi: 10.1007/s10681-013-0926-y
- Gnirke A, Melnikov A, Maguire J, et al (2009) Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat Biotechnol* 27:182–189. doi: 10.1038/nbt.1523
- Gomez C, Dussert S, Hamon P, et al (2009) Current genetic differentiation of *Coffea canephora*

- Pierre ex A. Froehn in the Guineo-Congolian African zone: cumulative impact of ancient climatic changes and recent human activities. *BMC Evol Biol* 9:1–19. doi: 10.1186/1471-2148-9-167
- Grandillo S (2014) Introgression Libraries with Wild Relatives of Crops. *Genomics Plant Genet Resour* 2:87–122. doi: 10.1007/978-94-007-7575-6_4
- Kosman E, Leonard KJ (2005) Similarity coefficients for molecular markers in studies of genetic relationships between individuals for haploid, diploid, and polyploid species. *Mol Ecol* 14:415–424. doi: 10.1111/j.1365-294X.2005.02416.x
- Kumar S, Stecher G, Tamura K (2016) MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol* 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.). *Theor Appl Genet* 100:139–146.
- Leroy T, Marraccini P, Dufour M, et al (2005) Construction and characterization of a *Coffea canephora* BAC library to study the organization of sucrose biosynthesis genes. *Theor Appl Genet* 111:1032–1041. doi: 10.1007/s00122-005-0018-z
- Liao P-Y, Lee KH (2010) From SNPs to functional polymorphism: The insight into biotechnology applications. *Biochem Eng J* 49:149–158. doi: 10.1016/j.bej.2009.12.021
- Marraccini P, Vinecky F, Alves GSC, et al (2012) Differentially expressed genes and proteins upon drought acclimation in tolerant and sensitive genotypes of *Coffea canephora*. *J Exp Bot* 63:4191–4212. doi: 10.1093/jxb/ers103
- Metsalu T, Vilo J (2015) ClustVis: a web tool for visualizing clustering of multivariate data using Principal Component Analysis and heatmap. *Nucleic Acids Res* 43:W566–W570. doi: 10.1093/nar/gkv468
- Musoli P, Cubry P, Aluka P, et al (2009) Genetic differentiation of wild and cultivated populations: diversity of *Coffea canephora* Pierre in Uganda. *Genome* 52:634–646. doi: 10.1139/G09-037
- Neves LG, Davis JM, Barbazuk WB, Kirst M (2013) Whole-exome targeted sequencing of the uncharacterized pine genome. *Plant J* 75:146–156. doi: 10.1111/tpj.12193
- Neves LG, Davis JM, Barbazuk WB, Kirst M (2014) A High-Density Gene Map of Loblolly Pine (*Pinus taeda* L.) Based on Exome Sequence Capture Genotyping. *G3 Genes|Genomes|Genetics* 4:29–37. doi: 10.1534/g3.113.008714
- Ojopi EPB, Gregorio SP, Guimarães PEM, et al (2004) O genoma humano e as perspectivas para o estudo da esquizofrenia. *Rev Psiquiatr Clínica* 31:9–18. doi: 10.1590/S0101-60832004000100003

- Pinto LA, Stein RT, Kabesch M (2008) Impact of genetics in childhood asthma. *J Pediatr (Rio J)* 84:S68-75. doi: 10.2223/JPED.1781
- Prakash NS, Combes M-C, Dussert S, et al (2005) Analysis of genetic diversity in Indian robusta coffee gene pool (*Coffea canephora*) in comparison with a representative core collection using SSRs and AFLPs. *Genet Resour Crop Evol* 52:333–343. doi: 10.1007/s10722-003-2125-5
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945–959. doi: 10.1111/j.1471-8286.2007.01758.x
- Ren J, Sun D, Chen L, et al (2013) Genetic Diversity Revealed by Single Nucleotide Polymorphism Markers in a Worldwide Germplasm Collection of Durum Wheat. *Int J Mol Sci* 14:7061–7088. doi: 10.3390/ijms14047061
- 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, California, p 1
- Resende M De, Lopes P, Silva RL Da, Pires IE (2008) Seleção genômica ampla (GWS) e maximização da eficiência do melhoramento genético. *Pesqui Florest Bras* 56:63–77. doi: 51970036
- Sera T, Ruas PM, Ruas CDF, et al (2003) Genetic polymorphism among 14 elite *Coffea arabica* L. cultivars using RAPD markers associated with restriction digestion. *Genet Mol Biol* 26:59–64. doi: 10.1590/S1415-47572003000100010
- Stacklies W, Redestig H, Scholz M, et al (2007) pcaMethods a bioconductor package providing PCA methods for incomplete data. *Bioinformatics* 23:1164–1167. doi: 10.1093/bioinformatics/btm069
- Vieira LGE, Andrade AC, Colombo CA, et al (2006) Brazilian coffee genome project: an EST-based genomic resource. *Brazilian J Plant Physiol* 18:95–108. doi: 10.1590/S1677-04202006000100008
- Yang W, Kang X, Yang Q, et al (2013) Review on the development of genotyping methods for assessing farm animal diversity. *J Anim Sci Biotechnol* 4:2. doi: 10.1186/2049-1891-4-2
- Zhang J, Song Q, Cregan PB, et al (2015) Genome-wide association study for flowering time, maturity dates and plant height in early maturing soybean (*Glycine max*) germplasm. *BMC Genomics* 16:1–11. doi: 10.1186/s12864-015-1441-4
- Zhang P, Li J, Li X, et al (2011) Population structure and genetic diversity in a rice core collection (*Oryza sativa* L.) investigated with SSR markers. *PLoS One* 6:e27565. doi: 10.1371/journal.pone.0027565

Zhou L, Vega FE, Tan H, et al (2016) Developing Single Nucleotide Polymorphism (SNP) Markers for the Identification of Coffee Germplasm. *Trop Plant Biol* 9:82–95. doi: 10.1007/s12042-016-9167-2

CAPÍTULO 2

**Ganhos com seleção e análise de diversidade genética em populações de *Coffea canephora*
via modelos mistos (REML/BLUP)**

RESUMO

A espécie *Coffea canephora* destaca-se por apresentar populações com expressiva variabilidade genética, as quais podem ser divididas em dois grupos heteróticos denominados café Conilon e café Robusta. Para o sucesso dos programas de melhoramento genético dessa espécie, esses grupos heteróticos podem ser explorados separadamente ou hibridizado, mas estratégias eficientes de melhoramento e seleção de plantas que permitam identificar genótipos superiores são imprescindíveis. A metodologia de modelos mistos (REML/BLUP) permite a predição acurada e não viesada dos valores genéticos e por meio destes, a seleção de genótipos superiores. O objetivo desse estudo foi estimar os parâmetros genéticos, quantificar os ganhos com a seleção e avaliar a diversidade genética de genótipos oriundos dos grupos varietais Conilon e Robusta e de famílias híbridas obtidas de cruzamentos entre Conilon e Robusta, utilizando a metodologia de modelos mistos (REML/BLUP). Características agronômicas foram avaliadas em 147 genótipos/famílias. As análises foram realizadas pelo *software* Selegen REML/BLUP. O índice de seleção permitiu ganhos expressivos com a seleção, em relação a população original, de 54,84%, 31,94% e 16,99%, nas populações de Conilon, Robusta e famílias híbridas, respectivamente. Os genótipos Conilon UFV 3628-45, UFV 3627-27, UFV 3628-22 e UFV 3629-34, bem como os genótipos Robusta UFV 3360-169, UFV 3630-5 e UFV 3368-58 se mostraram promissores por serem geneticamente divergentes e por proporcionarem ganhos elevados com a seleção, podendo ser inter cruzados para obter cultivares dentro de cada grupo varietal ou como genitores em cruzamentos interpopulacionais. As famílias H097, H0912, H0918 e H0921 se destacaram em relação as demais famílias pela diversidade e ganho com seleção. Além disso, as famílias H0912, H0918 e H0921 também apresentaram maiores ganhos com a seleção do que os genitores que as originaram, demonstrando a ocorrência de heterose. Dessa forma, estas quatro famílias podem ser inter cruzadas (cruzamento intrapopulacional) para avanço da seleção recorrente e também serem testadas em diferentes regiões para formar uma combinação de sementes de variedades híbridas.

Palavras-chave: Café, grupos heteróticos, seleção de famílias, diversidade genética.

Introdução

A cafeicultura é uma das atividades agrícolas de maior importância no mercado internacional, com o Brasil se destacando como o maior produtor e exportador mundial (Bicho et al. 2013). Dentro do gênero *Coffea*, que inclui mais de 100 espécies, as duas espécies cultivadas de importância comercial são *Coffea arabica* e *Coffea canephora* (Babova et al. 2016).

O cultivo comercial da espécie *C. arabica* teve um rápido desenvolvimento, especialmente devido a qualidade superior de bebida que essa espécie apresenta, em relação a *C. canephora*. No entanto, entre os anos de 1870 a 1900 com a constatação da grande incidência de ferrugem (*Hemileia vastatrix*) nas lavouras de café, a espécie *C. canephora* passou a ter ênfase entre os melhoristas dessa cultura, por apresentar resistência genética ao agente causal dessa doença (Charrier and Berthaud 1988). A partir daí essa espécie também tem se destacado nos programas de melhoramento por sua rusticidade, elevado potencial produtivo e maior conteúdo de sólidos solúveis totais, em comparação ao café arabica, o que tem despertado a atenção das indústrias de café. A espécie *C. canephora* se caracteriza por ser alógama, diploide ($2n=2x=22$), com sistema de autoincompatibilidade gametofítica e adaptada a regiões de baixas altitudes e com temperaturas mais elevadas (Leroy et al. 2005). Além disso, essa espécie se destaca por apresentar populações com expressiva variabilidade genética, as quais podem ser divididas em dois grupos heteróticos.

Os grupos heteróticos denominados de café Conilon (*Coffea canephora* var. Conilon) e café Robusta (*Coffea canephora* var. Robusta) apresentam características distintas e complementares e são de suma importância nos programas de melhoramento de *C. canephora* para exploração da heterose (vigor híbrido) (Musoli et al. 2009). Os cafeeiros denominados Conilon se destacam por serem mais tolerantes a seca e estarem presentes nas principais variedades desenvolvidas no Brasil. Os genótipos denominados Robusta compreendem plantas mais altas, vigorosas, com folhas e frutos maiores, resistentes à ferrugem do cafeeiro, porém são mais sensíveis à seca (Marraccini et al. 2012).

Em um programa de melhoramento genético a presença de variabilidade genética é condição básica e necessária para se obter sucesso (Babova et al. 2016). Associado a isso, estratégias de melhoramento e seleção de plantas, cada vez mais eficientes, que permitam identificar genótipos superiores são imprescindíveis. Para definir tais estratégias, é necessário o conhecimento de alguns aspectos biológicos específicos da cultura (Ferreira et al. 2005), bem como teste e adaptação de metodologias estatísticas

Espécies vegetais perenes como o cafeeiro apresentam longo ciclo reprodutivo; acentuada oscilação anual de produção (bienalidade); caracteres se expressando ao longo de vários anos e

diferenças na precocidade e longevidade produtiva (Sera 2001). Essas características resultam em dados experimentais desbalanceados e, conseqüentemente, com maior complexidade na estimação dos parâmetros genéticos. Essa limitação tem dificultado a identificação dos melhores genótipos nos programas de melhoramento genético do cafeeiro e, portanto, métodos estatísticos mais eficientes para a seleção se fazem necessários. Nesse sentido, os métodos tradicionais de estimação com base na análise de variância não são os mais recomendados para a análise de dados nos programas de melhoramento do cafeeiro (Oliveira et al. 2014; Resende et al. 2014).

O uso de procedimentos genético-estatísticos mais refinados e precisos na estimação de valores genéticos como a metodologia de modelos mistos do tipo REML/BLUP (*Residual or Restricted Maximum Likelihood/Best Linear Unbiased Prediction*) (Henderson 1975), tem se mostrado vantajoso em estudos de genética quantitativa e na seleção de plantas em espécies perenes (Costa et al. 2010; Maia et al. 2011; Oliveira et al. 2014; Resende et al. 2014; Spinelli et al. 2015; Carias et al. 2016). Essa metodologia permite a predição acurada e não viesada dos valores genéticos, mesmo sob condições de dados experimentais desbalanceados, propiciando maximização do ganho genético com a seleção; não exige que os dados sejam obtidos sob estruturas rígidas de experimentação; permite simultaneamente o uso das informações do indivíduo, família e de medidas repetidas no tempo, além da possibilidade de considerar um grande número de informações obtidas de vários experimentos e permitir a correção dos dados para efeitos ambientais (Henderson 1975; Resende 2002; Pereira et al. 2013). Com esse método, a partir de dados fenotípicos, os componentes de variância são obtidos por meio da metodologia REML e os valores genéticos preditos por meio do BLUP (Resende 2002).

O objetivo desse estudo foi estimar os parâmetros genéticos, quantificar os ganhos com a seleção e avaliar a diversidade genética de genótipos oriundos dos grupos varietais Conilon e Robusta e de famílias híbridas (Conilon x Robusta), utilizando a metodologia de modelos mistos (REML/BLUP), para características agronômicas avaliadas.

Material e métodos

Material genético

As populações de estudo foram compostas por clones dos grupos varietais Conilon e Robusta e por famílias híbridas originadas de cruzamentos entre estes grupos. O material genético de Conilon foi obtido do Instituto Capixaba de Pesquisa, Assistência Técnica e Extensão Rural (INCAPER) e o material de Robusta foi obtido do Centro Agronómico Tropical de Investigación y Enseñanza (CATIE). Esses materiais genéticos compõem o programa de melhoramento da

Empresa de Pesquisa Agropecuária de Minas Gerais (Epamig) em parceria com a Universidade Federal de Viçosa (UFV) e a Empresa Brasileira de Pesquisa Agropecuária - Café (Embrapa Café), localizado no município de Oratórios/MG.

As plantas foram dispostas em espaçamento de 3,0 m entre fileiras e 1,5 m entre plantas. O plantio dos genótipos dos grupos Conilon e Robusta foi realizado em julho de 2009 e o plantio das famílias foi realizado em março de 2011.

Os genótipos Conilon foram avaliados em dois grupos experimentais com testemunhas comuns (clones). O mesmo procedimento foi realizado para a população de Robusta. A utilização de dois experimentos por população se deve a preservação de variância comum dentro de cada bloco. Para Conilon foram utilizados 71 genótipos (Tabela 1) em delineamento em blocos casualizados (DBC) com 5 repetições e uma planta por parcela. As plantas das repetições correspondem a clones. Os dois grupos experimentais ficaram compostos por 34 genótipos/grupo e três testemunhas comuns/grupo (UFV 3629-11, UFV 3628-2 e UFV 513). Para Robusta foram utilizados 56 genótipos (Tabela 2) em DBC com 5 repetições (clones) e uma planta por parcela, divididos em dois grupos experimentais de 27 clones/grupo e duas testemunhas comuns/grupo (UFV 3366-134 e UFV 3366-139).

As 20 famílias híbridas que correspondem a híbridos interpopulacionais foram obtidas, em setembro de 2009, por meio de cruzamentos artificiais entre cinco genótipos do grupo Conilon (genitores masculinos) e cinco do grupo Robusta (genitores femininos), avaliados em um dialelo parcial interpopulacional (Tabela 3). O delineamento estatístico utilizado foi o de blocos casualizados com 1 a 35 plantas/família tidas como repetições e uma planta por parcela, sendo estas obtidas a partir de sementes de cada cruzamento.

Avaliações fenotípicas

Os genótipos foram avaliados para oito características fenotípicas. As avaliações foram realizadas em sete anos consecutivos (2010 a 2016) para os genótipos dos grupos varietais Conilon e Robusta e em cinco anos consecutivos (2012 a 2016) para as famílias híbridas. Estas foram realizadas na época de maturidade fisiológica dos frutos dos cafeeiros.

Foram avaliadas cinco características categóricas: vigor vegetativo (Vig), avaliação da incidência à ferrugem do cafeeiro no campo (Fer), incidência de cercosporiose (Cer), época de maturação dos frutos (Mat) e tamanho de fruto (TF). O vigor vegetativo foi avaliado pelo aspecto geral da planta, observando-se o enfolhamento, a coloração das folhas, o estado nutricional e a sanidade dos cafeeiros. Foi utilizada a escala de notas de 1 a 10, sendo a nota 1 atribuída às plantas

totalmente depauperada e a 10 às plantas altamente vigorosas; a incidência à ferrugem do cafeeiro, causada pelo fungo *Hemileia vastatrix* Berk. & Br., foi avaliada com notas de 1 a 5. A nota 1 foi atribuída para genótipos sem nenhum sintoma do patógeno e nota 5 para genótipos altamente suscetíveis; a incidência à cercospora no campo foi avaliada por notas de 1 a 5. A nota 1 foi atribuída para genótipos sem nenhum sintoma do patógeno e nota 5 para genótipos altamente suscetíveis; por ocasião da atribuição de notas ao vigor vegetativo, a época de maturação dos frutos do cafeeiro foi classificada em precoce, média e tardia, recebendo notas de 1 a 3 para esses atributos, respectivamente; e o tamanho dos frutos foi classificado como pequeno (nota 1), médio (nota 2) e graúdo (nota 3).

As características contínuas avaliadas foram: altura da planta (AP), diâmetro da projeção da copa (DC) e produção em litros por planta (Prod). A altura da planta foi determinada em centímetros (cm), pela medida da ramificação ortotrópica mais desenvolvida, do nível do solo até o último ponto apical do cafeeiro com o auxílio de uma trena métrica afixada em uma haste de madeira; o diâmetro da projeção da copa da planta foi determinado em centímetros (cm) por meio de régua no sentido perpendicular à linha de plantio; e a produção por planta de cafeeiros foi avaliada colhendo todos os frutos presentes em um genótipo e mensurado o volume total em litros de café fresco, recém colhidos.

Análise de Deviance (ANADEV)

A análise de deviance para os genótipos Conilon, Robusta e para as famílias híbridas foram realizadas por meio da estimativa de cada um dos efeitos do modelo estudado. Inicialmente obtêm-se os valores de deviance para o modelo completo (com todos os efeitos) e para cada um dos efeitos do modelo (sem o referido efeito de interesse). De posse dos valores de deviance foi obtida a razão de verossimilhança (LR) por meio da diferença entre as deviances do modelo sem o efeito a ser testado e o modelo completo. A significância dessa diferença foi testada via teste da razão de verossimilhança (TRV), através da tabela de Qui-quadrado com 1 grau de liberdade, por só testar um efeito (Parâmetro) por vez, a 1 e 5% de probabilidade.

Parâmetros genéticos dos genótipos Conilon e Robusta

As análises dos dados fenotípicos foram realizadas a partir da metodologia REML/BLUP, também denominada metodologia de modelos mistos (Henderson 1975), utilizando o *software* Selegen REML/BLUP (Resende 2016). O modelo utilizado foi: $y = Xu + Zg + Wb + Tp + e$, em que y é o vetor de dados; u é o vetor de efeitos das médias de anos (assumidos como fixos),

somados à média geral; g é o vetor de efeitos genotípicos de progênie (assumidos como aleatórios); b é o vetor dos efeitos dos blocos (assumidos como aleatórios); p é o vetor dos efeitos de ambiente permanente (assumidos como aleatórios); e é o vetor de resíduos (assumidos como aleatórios). As letras maiúsculas representam as matrizes de incidência para os referidos efeitos.

As estruturas de médias e variâncias associadas ao modelo são: $y|u, V \sim N(Xu, V)$; $g|G, \sigma_g^2 \sim N(0, G\sigma_g^2)$; $b|I\sigma_b^2 \sim N(0, I\sigma_b^2)$; $p|I\sigma_p^2 \sim N(0, I\sigma_p^2)$; $e|I\sigma_e^2 \sim N(0, I\sigma_e^2)$. As covariâncias entre os efeitos são iguais à zero.

Em que:

$$V = \text{Variância fenotípica e equivale a: } ZG^{-1}Z'\sigma_g^2 + WW'\sigma_b^2 + TT'\sigma_p^2 + I\sigma_e^2;$$

G = Matriz de correlação genotípica entre progênies em avaliação;

I = Matriz Identidade;

$\sigma_g^2, \sigma_b^2, \sigma_p^2$, e σ_e^2 são as variâncias genotípica entre progênies, entre bloco, de ambiente permanente, e residual, respectivamente.

As equações de modelos mistos para o BLUP (*Best Linear Unbiased Predictor*) equivalem

a:

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

Em que:

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

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

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

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

$$b^2 = \frac{\sigma_b^2}{\sigma_g^2 + \sigma_b^2 + \sigma_p^2 + \sigma_e^2}: \text{coeficiente de determinação dos efeitos entre blocos.}$$

$$p^2 = \frac{\sigma_p^2}{\sigma_g^2 + \sigma_b^2 + \sigma_p^2 + \sigma_e^2}: \text{coeficiente de determinação dos efeitos de ambiente permanente};$$

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 2007; Resende and Duarte 2007), cujos estimadores são:

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

$$\sigma_b^2 = [\hat{b}'\hat{b} + \sigma_e^2 \text{tr}(C^{33})]/r;$$

$$\sigma_p^2 = [\hat{p}'\hat{p} + \sigma_e^2 \text{tr}(C^{44})]/N;$$

$$\sigma_e^2 = [y'y - \hat{u}'X'y - \hat{g}Z'y - \hat{b}W'y - \hat{p}V'y]/[N - r(X)].$$

Em que:

G: Matriz de correlação genotípica entre progênes 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, r: número total de dados, de progênes, e de blocos, respectivamente.

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

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

Parâmetros genéticos das famílias híbridas

Para análise dos parâmetros genéticos foi utilizado o modelo: $y = Xu + Tc + Wf + Zm + Qs + Sb + e$, em que, y é o vetor de dados; u é o vetor de efeitos de médias de anos (assumidos como fixos) somados à média geral; c é o vetor de efeitos da capacidade específica de combinação entre os genitores Conilon e Robusta (assumidos como aleatórios); f é o vetor de efeitos da capacidade geral de combinação do parental Robusta (assumidos como aleatórios); m é o vetor de efeitos da capacidade geral de combinação do parental Conilon (assumidos como aleatórios); s é o vetor de efeitos de ambiente permanente de indivíduos (assumidos como aleatórios); b é o vetor de efeitos de ambiente permanente de blocos (assumidos como aleatórios); e é o vetor de resíduos (assumidos como aleatórios). As letras maiúsculas representam as matrizes de incidência para os referidos efeitos.

As estruturas de médias e variâncias associadas ao modelo são: $y|r, V \sim N(Xu, V)$; $m|\sigma_m^2 \sim N(0, I\sigma_m^2)$; $f|\sigma_f^2 \sim N(0, I\sigma_f^2)$; $c|\sigma_c^2 \sim N(0, I\sigma_c^2)$; $s|\sigma_s^2 \sim N(0, I\sigma_s^2)$; $b|\sigma_b^2 \sim N(0, I\sigma_b^2)$; $e|\sigma_e^2 \sim N(0, I\sigma_e^2)$. As covariâncias são entre os efeitos são iguais à zero.

Em que:

V = Variância fenotípica e equivale a: $ZZ'\sigma_m^2 + WW'\sigma_f^2 + TT'\sigma_c^2 + QQ'\sigma_s^2 + SS'\sigma_b^2 + I\sigma_e^2$;

I = Matriz Identidade;

$\sigma_m^2, \sigma_f^2, \sigma_c^2, \sigma_s^2, \sigma_b^2$, e σ_e^2 são as variâncias entre os genitores do grupo Conilon, entre os genitores do grupo Robusta, da capacidade específica de combinação entre os genitores dos grupos Conilon e Robusta, de ambiente permanente de indivíduo, de ambiente permanente de blocos, e residual, respectivamente.

As equações de modelos mistos para o BLUP (*Best Linear Unbiased Predictor*) equivalem

a:

$$\begin{bmatrix} \hat{u} \\ \hat{m} \\ \hat{f} \\ \hat{c} \\ \hat{s} \\ \hat{b} \end{bmatrix} = \begin{bmatrix} X'X & X'Z & X'W & X'T & X'Q & X'S \\ Z'X & Z'Z + I\lambda_1 & Z'W & Z'T & Z'Q & Z'S \\ W'X & W'Z & W'W + I\lambda_2 & W'T & W'Q & W'S \\ T'X & T'Z & T'W & T'T + I\lambda_3 & T'Q & T'S \\ Q'X & Q'Z & Q'W & Q'T & Q'Q + I\lambda_4 & Q'S \\ S'X & S'Z & S'W & S'T & S'Q & S'S + I\lambda_5 \end{bmatrix}^{-1} \begin{bmatrix} X'y \\ Z'y \\ W'y \\ T'y \\ Q'y \\ S'y \end{bmatrix}$$

Em que:

$$\lambda_1 = \sigma_e^2 / \sigma_m^2 = \frac{1-m^2-f^2-c^2-s^2-b^2}{m^2};$$

$$\lambda_2 = \sigma_e^2 / \sigma_f^2 = \frac{1-m^2-f^2-c^2-s^2-b^2}{f^2};$$

$$\lambda_3 = \sigma_e^2 / \sigma_c^2 = \frac{1-m^2-f^2-c^2-s^2-b^2}{c^2};$$

$$\lambda_4 = \sigma_e^2 / \sigma_s^2 = \frac{1-m^2-f^2-c^2-s^2-b^2}{s^2};$$

$$\lambda_5 = \sigma_e^2 / \sigma_b^2 = \frac{1-m^2-f^2-c^2-s^2-b^2}{b^2};$$

$m^2 = \frac{\sigma_m^2}{\sigma_m^2 + \sigma_f^2 + \sigma_c^2 + \sigma_s^2 + \sigma_b^2 + \sigma_e^2}$: coeficiente de determinação da capacidade geral de combinação do parental Conilon;

$f^2 = \frac{\sigma_f^2}{\sigma_m^2 + \sigma_f^2 + \sigma_c^2 + \sigma_s^2 + \sigma_b^2 + \sigma_e^2}$: coeficiente de determinação da capacidade geral de combinação do parental robusta;

$c^2 = \frac{\sigma_g^2}{\sigma_m^2 + \sigma_f^2 + \sigma_c^2 + \sigma_s^2 + \sigma_b^2 + \sigma_e^2}$: coeficiente de determinação da capacidade específica de combinação entre os genitores Conilon e robusta;

$s^2 = \frac{\sigma_p^2}{\sigma_m^2 + \sigma_f^2 + \sigma_c^2 + \sigma_s^2 + \sigma_b^2 + \sigma_e^2}$: coeficiente de determinação de ambiente permanente de indivíduos;

$b^2 = \frac{\sigma_b^2}{\sigma_m^2 + \sigma_f^2 + \sigma_c^2 + \sigma_s^2 + \sigma_b^2 + \sigma_e^2}$: coeficiente de determinação de ambiente permanente de blocos.

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

$$\sigma_m^2 = [\hat{m}'\hat{m} + \sigma_e^2 \text{tr}(C^{22})]/q;$$

$$\sigma_f^2 = [\hat{f}'\hat{f} + \sigma_e^2 \text{tr}(C^{33})]/n;$$

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

$$\sigma_s^2 = [\hat{s}'\hat{s} + \sigma_e^2 \text{tr}(C^{55})]/N;$$

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

$$\sigma_e^2 = [y'y - \hat{u}'X'y - \hat{m}Z'y - \hat{f}W'y - \hat{c}T'y - \hat{s}Q'y - \hat{b}S'y]/[N - r(X)].$$

Em que:

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, n, m e l: número total de dados, de genitores do grupo Conilon, genitores do grupo Robusta, número de combinações entre os grupos Conilon e Robusta; e de blocos, respectivamente.

O coeficiente de variação genética (CV_g) e o coeficiente de variação residual foram calculados conforme citado anteriormente.

Correlação genética

A correlação genética entre as variáveis avaliadas para os genótipos Conilon, Robusta e para as famílias híbridas foi estimada de acordo com (Falconer and Mackay 1996): $r_g = \frac{\hat{\sigma}_{gxy}}{\sqrt{\hat{\sigma}_{gx}^2 \hat{\sigma}_{gy}^2}}$,

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.

Índice de seleção

A identificação dos genótipos superiores de Conilon, Robusta e das famílias híbridas foi realizada pelo índice de *rank* médio de (Mulamba and Mock 1978) adaptado por (Resende 2007). Este índice consiste em classificar os genótipos em ordem favorável ao melhoramento para cada característica analisada e, após a classificação, é calculada a média dos ranks. Com base no resultado deste índice, foram estimados os ganhos com a seleção.

Diversidade genética

A dissimilaridade genética foi obtida por meio da distância euclidiana média e o agrupamento foi realizado pelo método de agrupamento de Tocher (Rao 1952), o qual se baseia na formação de grupos cujas distâncias dentro dos grupos sejam menores que as distâncias entre grupos.

Análises conjunta de índice de seleção

Com o intuito de avaliar a ocorrência de heterose nas famílias híbridas, além das análises realizadas em cada população (Conilon, Robusta e famílias híbridas), foi realizada análise conjunta de índice de rank médio (Mulamba and Mock 1978), para as famílias e seus respectivos genitores, Conilon e Robusta (Tabela 3). Essa análise foi realizada com base em três anos de avaliações (2014 a 2016). O genitor Robusta UFV 3374-28 não foi incluído na análise pelo fato do mesmo não ter sido fenotipado.

Seleção de Híbridos para as variáveis incidência à ferrugem e produção por planta

Foi realizada uma análise para seleção dos melhores Híbridos dentro das famílias híbridas, para as variáveis incidência à ferrugem do cafeeiro e produção por planta. Essa análise foi realizada utilizando o modelo 170 do *software* Selegen REML/BLUP (Resende 2016).

Resultados e discussão

Análises dos genótipos do grupo varietal Conilon

A análise de deviance (ANADEV) para as variáveis analisadas na população do grupo varietal Conilon está apresentada na Tabela 4. Todas as variáveis analisadas apresentaram resultados significativos para a fonte de variação genotípica, pelo teste da razão de verossimilhança (TRV), indicando a existência de variabilidade genética entre os genótipos Conilon e, conseqüentemente, a possibilidade de obter ganhos com a seleção. Em trabalho realizado com

clones do grupo Conilon, utilizando a metodologia de modelos mistos, também foi constatada expressiva variabilidade genética entre os genótipos, o que favoreceu a obtenção de ganhos com a seleção clonal (Ramalho et al. 2016). Para a fonte de variação ambiental entre blocos, todas as variáveis, com exceção da característica época de maturação dos frutos apresentaram resultados significativos. Esse resultado evidencia a adoção correta do delineamento estatístico utilizado (DBC), pois foi verificada variação significativa entre os blocos, a qual foi corretamente controlada. A escolha correta do delineamento estatístico é imprescindível na precisão experimental, pois esta permite a redução da variância residual. Todas as variáveis apresentaram resultados significativos para a fonte de variação ambiente permanente, exceto as variáveis incidência de cercosporiose, maturação dos frutos e tamanho do fruto.

As estimativas dos parâmetros genéticos para as variáveis analisadas estão apresentadas na Tabela 5. Os parâmetros genéticos estimados possibilitam o entendimento sobre a estrutura e o potencial genético da população e conseqüentemente, permitem a seleção mais eficiente e acurada dos indivíduos em uma população.

A variância genética (V_g) assumiu valores entre 0,01 (Cer) a 26,45 (AP) e a variância residual (V_e) valores entre 0,11 (Cer) a 485,08 (DC). Para todas as variáveis analisadas as estimativas de V_g foram inferiores as estimativas de V_e , dessa forma os valores de V_g são considerados de baixa magnitude. Estes resultados indicam baixa variabilidade genética entre os genótipos Conilon, o que pode ser justificada pelo fato dos genótipos avaliados pertencerem a três famílias de meios-irmãos (UFV 3627, UFV 3628 e UFV 3629). Carias et al. (2016) avaliando oito progênies de Conilon meios-irmãos para as variáveis tamanho dos grãos, produtividade, porte, vigor vegetativo e grau de inclinação, verificaram, para a maioria dessas variáveis, valores de variância genética considerados de baixa magnitude. Os autores justificam que possivelmente os baixos valores observados podem ser atribuídos ao maior grau de parentesco entre as progênies avaliadas (Carias et al. 2016).

A herdabilidade individual no sentido amplo (h^2_g), ou seja, considerando os efeitos aditivos e de dominância, para as variáveis avaliadas assumiram valores entre 0,03 (DC) e 0,25 (TF). O conhecimento da magnitude da herdabilidade é de suma importância nos programas de melhoramento porque ela determina o grau de dificuldade no melhoramento de um caráter, visto que quanto maior os valores de h^2_g , maiores os efeitos genéticos no fenótipo do indivíduo (Resende et al. 2004).

O coeficiente de repetibilidade (r) é obtido por meio de repetidas medições de uma característica em um mesmo indivíduo, fato que ocorre com frequência em espécies perenes como

o café. O coeficiente de repetibilidade apresenta importância fundamental na determinação do número necessário de medições por indivíduo ao longo de vários anos, o que permite prever o seu valor real com certo grau de confiabilidade. A repetibilidade pode ser classificada em alta ($r > 0,60$), média ($0,30 < r < 0,60$) e baixa ($r < 0,30$) (Resende 2002). Os valores de r variaram de 0,09 (Cer) a 0,55 (AP), sendo considerados de baixa a média magnitude. Maiores valores de coeficiente de repetibilidade permitem reduzir o número de medições repetidas em um mesmo indivíduo e com isso os programas de melhoramento serão otimizados e terão custos menores (Cruz et al. 2004).

O coeficiente de determinação dos efeitos de ambiente permanente (c^2_{perm}) fornece a variação ambiental que permanece de um ano para o outro (Mrode 2014; de Resende et al. 2015). Na análise de deviance foram verificados efeitos significativos para a fonte de variação ambiente permanente, exceto para as variáveis incidência de cercosporiose (Cer) e tamanho do fruto (TF) (Tabela 4). Os valores de c^2_{perm} variaram de 0,01 (Cer) a 0,38 (AP) (Tabela 5).

O coeficiente de determinação dos efeitos ambientais entre blocos (c^2_{bloc}) também é uma medida de qualidade experimental e indica a capacidade de teste, ou seja, a oportunidade dos genótipos experimentarem diferentes ambientes no mesmo experimento. A significância para a fonte de variação ambiental entre blocos, exceto para a característica Mat (Tabela 4), associada aos baixos valores de c^2_{bloc} , entre 1% (Mat) a 13% (AP) (Tabela 5), indicam que o delineamento utilizado foi eficiente e a capacidade do teste foi adequada (Resende 2002).

O coeficiente de variação genotípico (CV_{gi}) permite inferir sobre a quantificação da variabilidade genética presente na população em relação à média do caráter e, portanto, valores maiores são desejáveis. O parâmetro CV_{gi} apresentou valores entre 3,51% (AP) e 16,27% (Mat), indicando uma situação favorável para o processo de seleção das variáveis analisadas.

O coeficiente de variação residual (CV_e), o qual indica a precisão do experimento, assumiu valores entre 11,19% (AP) e 58,32% (Prod). De acordo com (Pimentel-Gomes 2009), valores de CV_e menores que 10% são considerados de baixa magnitude; CV_e entre 10% a 20% são considerado médios; CV_e entre 20% e 30% são considerados altos e valores acima de 30% são considerados muito altos. Levando-se em consideração essa classificação, os valores de CV_e para as variáveis incidência à ferrugem do cafeeiro (35,79%), época de maturação dos frutos (36,65%) e produção por planta (58,32%) são considerados de magnitude muito alta, indicando para estas variáveis uma maior variação causada pelo ambiente.

Outro parâmetro genético de importância em estudos de melhoramento de plantas é o $CV_{relativo}$, expresso pela relação entre CV_{gi} e CV_e . Os valores de $CV_{relativo}$ variaram entre 0,21 (DC)

e 0,59 (TF). Maiores valores desse parâmetro indicam que as variáveis analisadas apresentam maior controle genético e menor influência do ambiente (Vencovsky and Barriga 1992), fato observado no presente estudo, visto que as características que apresentaram maiores valores de coeficiente de variação relativo, Fer (0,42), Mat (0,44) e TF (0,59), apresentaram maiores valores de h^2_g .

No contexto da avaliação genotípica, o parâmetro estatístico considerado mais importante para avaliação da qualidade experimental e do valor genético predito do indivíduo é a acurácia seletiva (r_{gg}) (Pompeu Junior et al. 2013). A acurácia reflete a quantidade e qualidade das informações e procedimentos utilizados na predição dos valores genéticos dos indivíduos e é dependente da herdabilidade e da repetibilidade da característica analisada. Ela se baseia na correlação entre o valor genotípico verdadeiro do material genético e aquele estimado ou predito a partir das informações experimentais de campo dos indivíduos (Resende and Duarte 2007). A acurácia é classificada em muito alta, quando maior que 90%; alta, quando entre 70% e 90%; moderada, entre 50% e 70% e baixa, quando menor que 50% (Resende and Duarte 2007). As variáveis analisadas apresentaram acurácia seletiva moderada a alta, variando entre 53% (DC) a 89% (TF). As variáveis que apresentaram maiores valores de h^2_g , 0,14 (Fer), 0,16 (Mat) e 0,25 (TF), e maiores valores de $CV_{relativo}$, 0,42 (Fer), 0,44 (Mat) e 0,59 (TF), também foram as variáveis com maiores valores de acurácia seletiva, 0,86 (Fer), 0,84 (Mat) e 0,89 (TF). Para as variáveis com estimativa de acurácia seletiva superior a 70% (Vig, Fer, Cer, Mat e TF) a predição dos valores genéticos apresentam maior confiabilidade. Para a variável produção por planta o valor obtido de acurácia foi de 67% sendo, portanto, considerada de magnitude moderada. Resultados semelhantes foram encontrados por Ramalho et al. (2016), onde foi verificado valor de acurácia para essa variável de 86% em clones de Conilon.

A análise de correlação genética entre as variáveis analisadas está apresentada na Tabela 6. A correlação genotípica mede o grau de associação de natureza genética entre duas variáveis. Correlações genéticas altas entre dois caracteres indicam que, a alteração em um caráter, via seleção, promove alterações no outro. A correlação genética pode ser positiva, quando é possível obter ganhos para uma variável selecionando outra, ou negativa, quando a seleção em uma variável irá contribuir negativamente em outra (Resende 2002). Essa estratégia permite obter rápidos progressos nos programas de melhoramento genético, via seleção indireta, por permitir a seleção de variáveis que possuem herança complexa, de difícil mensuração e até mesmo a seleção precoce em espécies perenes, por meio de outra variável correlacionada e que seja de mais fácil mensuração (Oliveira et al. 2011; Verardi et al. 2011).

As principais causas da correlação genética entre variáveis são decorrentes da ligação gênica, de causa transitória, e pleiotropia (um loco influenciando mais de um caráter simultaneamente), de causa principal e permanente (Falconer and Mackay 1996; Resende 2015).

Os maiores valores de correlação genotípica negativa foram observados entre as variáveis vigor vegetativo e incidência de cercosporiose (-0,55) e entre as variáveis produção por planta e incidência de cercosporiose (-0,57). Levando-se em consideração o sentido e a magnitude dessas correlações, a seleção para essas variáveis seria vantajosa nos programas de melhoramento do cafeeiro, uma vez que se busca obter plantas com menor incidência de cercosporiose e que conseqüentemente, isso resulte em plantas mais vigorosas e mais produtivas.

Os maiores valores de correlação genotípica positiva foram observados entre as variáveis vigor vegetativo e altura da planta (0,58), vigor vegetativo e diâmetro da projeção da copa (0,69), incidência à ferrugem do cafeeiro e incidência de cercosporiose (0,57) e entre altura da planta e diâmetro da projeção da copa (0,68). Visto que o objetivo dos programas de melhoramento de *C. canephora* é obter plantas mais vigorosas e com maior diâmetro da projeção da copa, a seleção de uma dessas variáveis promoveria alterações positivas na outra correlacionada, por meio da seleção indireta. O mesmo é verificado para as variáveis incidência à ferrugem do cafeeiro e incidência de cercosporiose, uma vez que é interessante e vantajoso plantas com menores níveis de incidência dessas doenças. As dificuldades com a seleção serão observadas quando esta for realizada nas variáveis vigor vegetativo e altura de planta e nas variáveis diâmetro da projeção da copa e altura de planta. Isso ocorre quando duas variáveis estão correlacionadas positivamente e uma delas é indesejada. Para os programas de melhoramento o ideal é que as plantas sejam mais vigorosas e possuam um maior diâmetro da projeção da copa, entretanto para a variável altura de planta a seleção ideal é no sentido de reduzir a altura.

Para auxiliar na escolha dos materiais superiores foram utilizados índices de seleção com base na média de *ranks* (Mulamba and Mock 1978; Resende 2007). Essa seleção é baseada em diversas variáveis de interesse, o que resulta em maiores ganhos simultâneos (Freitas et al. 2012). O índice Mulamba e Mock se mostrou mais eficiente na seleção entre famílias de meios-irmãos de café Conilon em trabalho comparando a eficiência em ganhos com seleção obtidos com o índice Multiplicativo, índice Aditivo e o índice de Mulamba e Mock (Carias et al. 2016).

Os valores da Tabela 7 representam a posição (rank) que os 15 genótipos superiores (intensidade de seleção de 21%) assumiram para cada variável analisada e os ganhos com a seleção simultânea. Os dados de índice de seleção mostraram que com essa intensidade de seleção espera-se obter ganhos de 54,84% em relação a população original. Isso indica que é possível promover

o aumento simultâneo da concentração de alelos favoráveis na população (Pereira et al. 2013). Em trabalho conduzido com clones do grupo Conilon, considerando quatro safras e com intensidade de seleção de 10%, foi observado ganhos na produção de café beneficiado de 57,85% (Ramalho et al. 2016).

O genótipo UFV 3628-2 se destacou pelo maior ganho simultâneo (136,07%), considerando as oito variáveis. Esse alto desempenho pode estar relacionado ao fato desse genótipo ser um possível híbrido entre os dois grupos varietais, Conilon e Robusta. Em trabalho prévio (capítulo 1), foram utilizados marcadores moleculares SNP para análise de diversidade e estrutura genética de alguns genótipos Conilon. Nesse trabalho, as análises dos marcadores pelas metodologias *Unweighted Pair Group Method using Arithmetic averages* (UPGMA), algoritmo *Neighbor Joining* (NJ) e abordagem bayesiana sugeriram que o genótipo Conilon UFV 3628-2 é provavelmente resultante de cruzamento natural e está classificado de forma errônea. Além disso, esse genótipo se destacou pela sua produtividade, resistência à ferrugem e precocidade (dados apresentados no capítulo 1). Aumento de produtividade pode ser resultante de vigor híbrido e genótipos Conilon normalmente são suscetíveis à ferrugem dessa forma, a resistência a essa doença normalmente é obtida de Robusta (Cubry et al. 2008; Musoli et al. 2009).

Foi realizada ainda, uma análise de agrupamento multivariada dos genótipos Conilon, para todas as variáveis analisadas simultaneamente (Tabela 8). O agrupamento foi realizado pelo método de Tocher, baseado na distância genética euclidiana média, a qual separou os 71 genótipos em dois grupos distintos. A maioria dos genótipos Conilon (70) foram alocados em um único grupo, grupo 1. O genótipo UFV 3628-2 que se destacou pelo maior ganho simultâneo (136,07%) na análise de índice de seleção com base na média de ranks ficou alocado individualmente no segundo grupo. Esse resultado corrobora com a hipótese desse genótipo ser um híbrido natural entre Conilon e Robusta.

Com base nesses resultados, foi realizada uma nova análise de agrupamento multivariada dos genótipos Conilon, para todas as variáveis analisadas simultaneamente, desconsiderando da análise o genótipo UFV 3628-2, por se tratar de um híbrido natural (Tabela 9). Foi observada a formação de 15 grupos, o que indica diversidade genética entre os genótipos, considerando as variáveis avaliadas. Os genótipos UFV 3628-45, UFV 3627-27, UFV 3628-22 e UFV 3629-34 se destacam por estarem entre os nove genótipos com maiores ganhos simultâneos (Tabela 7) e por terem sido alocados em grupos distintos (Tabela 9). Portanto, estes genótipos devem ser priorizados nos programas de melhoramento para obter cultivares dentro do grupo varietal ou como genitores em cruzamentos interpopulacionais para exploração do vigor híbrido (Ferrão et al.

2013). Resultados similares foram encontrados por Silva et al. (2015), utilizando o método de Tocher, baseado na distância genética de Mahalanobis, estimados a partir de seis variáveis fenotípicas avaliadas nos anos de 2010 e 2011. Nesse trabalho 71 genótipos Conilon foram agrupados em 13 grupos distintos, indicando a presença de variabilidade genética na população (Silva et al. 2015).

Análises dos genótipos do grupo varietal Robusta

A análise de deviance (ANADEV) para as variáveis analisadas na população do grupo varietal Robusta está apresentada na Tabela 10. Todas as variáveis analisadas, exceto a variável diâmetro da projeção da copa, apresentaram resultados significativos para a fonte de variação genotípica pelo teste da razão de verossimilhança (TRV), indicando a existência de variabilidade genética entre os genótipos Robusta e, conseqüentemente, a possibilidade de obter ganhos com a seleção. Apenas as variáveis diâmetro da projeção da copa e época de maturação dos frutos apresentaram resultados significativos para a fonte de variação ambiental entre blocos. Para a fonte de variação ambiente permanente, as variáveis vigor vegetativo, incidência de cercosporiose, altura da planta e diâmetro da projeção da copa apresentaram resultados significativos pelo teste da razão de verossimilhança (TRV).

As estimativas dos parâmetros genéticos para as variáveis analisadas estão apresentadas na Tabela 11.

Os valores de variância genética (V_g) variaram entre 0,01 (Fer) a 73,88 (AP) e os valores de variância ambiental (V_e) entre 0,36 (Mat) a 657,64 (AP). Assim, embora a análise de deviance (Tabela 9) tenha mostrado resultados significativos para a fonte de variação genotípica pelo teste da razão de verossimilhança (TRV), para algumas dessas variáveis os valores de V_g foram considerados baixos. Para todas as variáveis, os valores de V_g foram inferiores as estimativas dos valores de V_e , indicando baixa variabilidade genética na população de Robusta.

O coeficiente de repetibilidade (r) assumiu valores entre 0,08 (Fer) a 0,40 (AP). Esses valores classificam a repetibilidade das variáveis em baixa a média magnitude (Resende 2002). Os maiores valores de r foram observados para as variáveis altura da planta (0,40) e diâmetro da projeção da copa (0,37). Quanto maior o valor do coeficiente de repetibilidade menor será o número necessário de medidas repetidas em um mesmo indivíduo. Estudando a repetibilidade em genótipos de café Robusta, para a variável produção, foi verificado que quatro colheitas seriam o suficiente para selecionar genótipos promissores com 80% de precisão, tornando-se inalterada a partir da sexta colheita (Fonseca et al. 2004).

O coeficiente de determinação dos efeitos de ambiente permanente (c^2_{perm}) fornece a variação ambiental que permanece de um ano para o outro e é uma informação auxiliar para determinar se o modelo proposto é adequado ou não (Mrode 2014; Resende 2015). Na análise de deviance verificou-se efeitos significativos para a fonte de variação ambiente permanente, para as variáveis vigor vegetativo, incidência de cercosporiose, altura da planta e diâmetro da projeção da copa (Tabela 9). Os valores de c^2_{perm} variaram de 0% (Fer) a 32% (AP).

O coeficiente de determinação dos efeitos ambientais entre blocos (c^2_{bloc}) indica a heterogeneidade ambiental entre as parcelas dentro do bloco. A significância para a fonte de variação ambiental entre blocos, observada para as variáveis diâmetro da projeção da copa e época de maturação dos frutos (Tabela 9), associada aos baixos valores de c^2_{bloc} , entre 0% (Fer e TF) e 5% (Mat) (Tabela 10), indicam que o delineamento utilizado foi eficiente e a capacidade do teste precisa (Resende 2002). Estimativas significativas dos efeitos de blocos associados a baixos c^2_{bloc} , entre 2 e 5%, também foram obtidos em população de café Robusta introduzida da Costa Rica, indicando boa homogeneidade entre parcelas dentro dos blocos (Mistro 2013). O que pode ter favorecido esses resultados é o fato de que no presente trabalho, bem como no trabalho desenvolvido por Mistro (2013) foram utilizados uma única planta por parcela. O uso de parcelas contendo um único indivíduo e com maior número de repetições melhoram as análises estatísticas em espécies perenes, devido o maior espaçamento utilizado entre as plantas poder proporcionar heterogeneidade dentro das parcelas (Resende 2002).

O parâmetro CV_{gi} apresentou valores entre 4,96% (Fer) e 21,56% (AP). Esses resultados indicam que há variabilidade genética na população para as variáveis analisadas. A presença de variabilidade genética é condição indispensável para realização da seleção (Falconer and Mackay 1996; Grandillo 2014).

O coeficiente de variação residual (CV_e), o qual indica a precisão do experimento, assumiu valores entre 16,18% (AP) e 81,40% (Prod). De acordo com (Pimentel-Gomes 2009), os valores de CV_e para as variáveis altura da planta (16,18%), diâmetro da projeção da copa (18,20%) e tamanho de fruto (16,27%) são considerados de média magnitude. Os valores de CV_e para as variáveis vigor vegetativo (21,02%), incidência à ferrugem do cafeeiro (27,85%), incidência de cercosporiose (20,22%) e época de maturação dos frutos (23,59%) são considerados de alta magnitude. Somente para variável produção por planta o valor de CV_e (81,40%) foi considerado muito alto. Para contornar os problemas inerentes à baixa precisão experimental e aumentar o sucesso na obtenção de ganhos com a seleção é necessário reduzir a variância ambiental ou aumentar a variância genética. A variância ambiental pode ser reduzida pela uniformização do

ambiente experimental e a variância genética é elevada quando se aumenta a variabilidade genética da população (Pompeu Junior et al. 2013).

O CV_{relativo} apresentou valores entre 0,26 (Prod) a 0,65 (TF). Os maiores valores de CV_{relativo} foram encontrados para as variáveis época de maturação dos frutos (0,51) e tamanho dos frutos (0,65). Os maiores valores de h^2_g foram observados também para as variáveis maturação dos frutos (0,19) e tamanho de fruto (0,29), bem como os maiores valores de acurácia seletiva (r_{gg}), os quais foram de 82% (Mat) e 87% (TF). Esses resultados determinam maior controle genético e menor influência ambiental para tais variáveis.

As variáveis analisadas apresentaram acurácia seletiva moderada a alta, variando entre 53% (DC) a 87% (TF) (Resende and Duarte 2007). Os valores de acurácia para as variáveis incidência à ferrugem do cafeeiro (75%), incidência de cercosporiose (75%), tamanho de fruto (87%) e maturação dos frutos (82%), indicam que a predição dos valores genéticos apresenta maior confiabilidade.

A análise de correlação genética entre as variáveis analisadas está apresentada na Tabela 12. Os maiores valores de correlação genotípica positiva foram obtidos entre as variáveis vigor vegetativo e altura da planta (0,63), vigor vegetativo e diâmetro da projeção da copa (0,80) e entre altura da planta e diâmetro da projeção da copa (0,68). Esses resultados também foram verificados na população de Conilon, o que confirma que essas variáveis apresentam forte correlação. A correlação entre as variáveis vigor vegetativo e diâmetro da projeção da copa (0,80) pode ser explorada de forma vantajosa uma vez que, o ideal é que as plantas sejam vigorosas e com maior diâmetro da projeção da copa. As dificuldades com a seleção, assim como na população de Conilon, serão quando esta for realizada nas variáveis vigor vegetativo e altura de planta e nas variáveis diâmetro da projeção da copa e altura de planta visto que, essas variáveis estão correlacionadas positivamente, mas a seleção para a variável altura de planta é desejável que seja no sentido de reduzir a altura das mesmas.

A variável diâmetro da projeção da copa (DC) não foi incluída nas análises subsequentes de índice de seleção de Mulamba-Mock e análise de agrupamento de Tocher, por não ter apresentado efeito genético significativo, de acordo com a análise de deviance (Tabela 10).

Na Tabela 13 estão representados os 15 melhores genótipos Robusta, os quais apresentaram maiores ganhos simultâneos para as sete variáveis com efeito genético significativo pela análise de deviance. O índice foi obtido com base na média de ranks (Mulamba and Mock 1978; Resende 2007). Nesse caso, em que foi utilizada uma intensidade de seleção em torno de 27% (15 genótipos), o ganho estimado foi de 31,94% em relação à população original. O valor do ganho

estimado foi expressivo, indicando o sucesso com a seleção simultânea para as variáveis analisadas.

Pela análise de agrupamento multivariada dos genótipos Robusta realizada pelo método de Tocher, baseado na distância genética euclidiana média, os 56 genótipos foram alocados em quatro grupos distintos (Tabela 14). No primeiro grupo foram alocados a maioria dos genótipos (52), no segundo grupo ficaram os genótipos UFV 3360-169 e UFV 3630-7 e no terceiro e quarto grupo foram alocados os genótipos UFV 3630-5 e UFV 3368-58, respectivamente. Em trabalho conduzido com 56 genótipos do grupo Robusta, avaliados para seis variáveis fenotípicas nos anos de 2010 e 2011, foi verificado pela análise de agrupamento de Tocher, baseado na distância genética de Mahalanobis, a formação de cinco grupos. Os resultados mostram que 51 dos 56 genótipos Robusta avaliados ficaram alocados em um único grupo (Silva et al. 2015).

Os genótipos UFV 3360-169, UFV 3630-5 e UFV 3368-58, além de estarem entre os 15 genótipos com maiores ganhos simultâneos (Tabela 13), se destacaram por terem sido alocados em grupos distintos (Tabela 14). Dessa forma, estes genótipos devem ser priorizados nos programas de melhoramento para obter cultivares dentro do grupo varietal Robusta ou como genitores em cruzamentos interpopulacionais para exploração do vigor híbrido (Ferrão et al. 2013).

Análises das famílias híbridas

A análise de deviance (ANADEV) para as variáveis analisadas nas famílias híbridas está apresentada na Tabela 15. As variáveis vigor vegetativo (Vig) e altura da planta (AP) apresentaram resultados significativos para a fonte de variação capacidade específica de combinação dos genitores da população de fêmeas com os genitores da população de machos, pelo teste da razão de verossimilhança (TRV). As variáveis incidência à ferrugem do cafeeiro (Fer) e época de maturação dos frutos (Mat) foram significativas para as fontes de variação genitores da população de fêmeas e genitores da população de machos, respectivamente. Esses resultados indicam a existência de variabilidade genética dentro da população de estudo para essas variáveis e, conseqüentemente, a possibilidade de obter ganhos com a seleção. Para a fonte de variação permanente de indivíduos, apenas as variáveis tamanho de fruto (TF) e produção por planta (Prod) não apresentaram resultados significativos pelo teste da razão de verossimilhança (TRV). A variável produção por planta (Prod) foi a única variável significativa pelo teste da razão de verossimilhança (TRV) para a fonte de variação blocos.

As estimativas dos parâmetros genéticos para as variáveis analisadas encontram-se descritas abaixo na Tabela 16.

Os valores de variância genética entre machos (V_{gm}), variância genética entre fêmeas (V_{gf}) e da variância da capacidade específica de combinação interpopulacional (V_{fm}) foram, de modo geral, baixos. A V_{gm} e a V_{gf} apresentaram valores expressivos apenas para as variáveis altura da planta (69,72) e diâmetro da projeção da copa (19,85), respectivamente. Para a V_{fm} , valores expressivos foram observados para duas variáveis, altura da planta (67,49) e diâmetro da projeção da copa (45,08). Dessa forma, é possível verificar a presença de variabilidade para as variáveis altura da planta e diâmetro da projeção da copa, na população das famílias de Híbridos analisadas e na população de genitores que deram origem a essas famílias Híbridas.

O coeficiente de determinação dos efeitos permanentes (c^2_{perm}) e o coeficiente de determinação dos efeitos de bloco (c^2_{bloc}), medidas de qualidade experimental, indicam a variação ambiental de um ano para outro e os valores genéticos não aditivos e a heterogeneidade ambiental entre as parcelas dentro do bloco, respectivamente. Os valores de c^2_{perm} variaram entre 1% (TF e Prod) e 31% (Vig). O c^2_{bloc} assumiram valores entre 1% (Fer, DC e Mat) a 4% (TF), indicando que o delineamento utilizado foi eficiente e a capacidade do teste foi adequada (Resende 2002).

Verifica-se que a repetibilidade (r) apresentou valores entre 0,05 (TF) a 0,47 (Vig), podendo ser classificada em baixa a média magnitude (Resende 2002). Como já discutido, quanto maior o valor de repetibilidade, menor o número de medições repetidas em um mesmo indivíduo serão necessários para predizer o seu valor real com certa confiabilidade. Os maiores valores de r foram verificados para as variáveis Vig (0,47), AP (0,46) e DC (0,35).

Com exceção de Fer e TF, todas as variáveis apresentaram dominância alélica em seus controles genéticos, conforme os valores de h^2_{dom} . Esse resultado revela que a heterose pode ser explorada no melhoramento dessas variáveis.

As variáveis analisadas apresentaram valores consideráveis de herdabilidade interpopulacional individual no sentido amplo, ou seja, dos efeitos genotípicos totais (h^2_g), exceto para a variável tamanho de fruto (0,02). O baixo valor de h^2_g para a variável TF resultou também em menor valor de $CV_{relativo}$ (0,01), indicando que essa variável apresenta menor controle genético e mais influência do ambiente. Portanto, as demais variáveis apresentam maior controle genético e, conseqüentemente, menos influência do ambiente, o que favorece o sucesso da seleção (Vencovsky and Barriga 1992).

O coeficiente de variação genotípico (CV_{gi}) apresentou valores considerados favoráveis à seleção das variáveis, com exceção da variável tamanho de fruto (0,25%), a qual apresentou também o menor valor de $CV_{relativo}$ (0,01) e de acurácia seletiva (0,05). Isso indica pouca existência de variabilidade genética para essa variável dentro da população das famílias de Híbridos.

Os valores de Coeficiente de variação residual (CV_e) foram considerados de magnitude média e muito alta (Pimentel-Gomes 2009), variando de 12,39% (AP) a 72,78% (Prod). O CV_e para as variáveis incidência à ferrugem do cafeeiro (52,72%), época de maturação dos frutos (40,47%) e produção por planta (72,78%), consideradas de alta magnitude, indica maior variação causada pelo ambiente.

O coeficiente de variação relativo (CV_{relativo}) apresentou valores entre 0,01 (TF) a 0,38 (AP). A acurácia seletiva (r_{gg}) variou entre 5% (TF) a 67% (AP), consideradas de baixa a moderada magnitude (Resende and Duarte 2007). É possível observar que os maiores valores de CV_{relativo} foram encontrados para as variáveis vigor vegetativo (0,33) e altura da planta (0,38). Os maiores valores de h^2_g foram observados também para as variáveis vigor vegetativo (0,39) e altura da planta (0,48), assim como os maiores valores de acurácia seletiva (r_{gg}) que foram 60% (Vig) e 67% (AP). Os resultados demonstram que as variáveis vigor vegetativo e altura da planta apresentaram maior proporção da variância genética explicando o caráter.

A análise de correlação genética entre as variáveis analisadas está apresentada na Tabela 17. O maior valor de correlação genotípica negativa foi observado entre as variáveis incidência à ferrugem do cafeeiro e época de maturação dos frutos (-0,50). Levando-se em consideração o sentido e a magnitude dessas correlações, a seleção em uma dessas variáveis não seria eficiente na seleção indireta da outra variável correlacionada. O ideal é que as variáveis incidência à ferrugem do cafeeiro e época de maturação dos frutos apresentassem uma alta correlação genotípica, mas que esta fosse positiva uma vez que, se busca obter plantas com menor incidência de ferrugem e que a maturação dos frutos seja mais precoce.

Os maiores valores de correlação genotípica positiva foram observados entre as variáveis vigor vegetativo e altura da planta (0,82), vigor vegetativo e diâmetro da projeção da copa (0,88), altura da planta e diâmetro da projeção da copa (0,86) e entre diâmetro da projeção da copa e produção por planta (0,44). Visto que o objetivo dos programas de melhoramento de *C. canephora* é obter plantas mais vigorosas, com maior diâmetro da projeção da copa e mais produtivas, a seleção de uma dessas variáveis promoveria alterações positivas na outra variável correlacionada, por meio da seleção indireta. A correlação observada entre as variáveis diâmetro da projeção da copa e produção por planta (0,44) é de suma importância para os programas de melhoramento do cafeeiro. A variável produção por planta (Prod) é uma característica quantitativa, governada por muitos genes e sofre grande influência do ambiente nesse sentido, ao selecionar plantas com maior diâmetro da projeção da copa, característica de mais fácil mensuração, indiretamente será praticada a seleção de plantas mais produtivas. Assim como na análise de correlação genotípica para os

genótipos Conilon e Robusta as dificuldades com a seleção para as famílias de Híbridos serão observadas quando esta for realizada nas variáveis vigor vegetativo e altura de planta e nas variáveis diâmetro da projeção da copa e altura de planta, pois elas estão correlacionadas positivamente, mas a seleção para altura de planta é realizada no sentido de reduzir a altura.

As variáveis diâmetro da projeção da copa (DC), tamanho de fruto (TF) e produção por planta (Prod) não apresentaram efeitos genéticos significativos, de acordo com a análise de deviance (Tabela 15) e, portanto, não foram incluídos nas análises subsequentes de índice de seleção de Mulamba-Mock e análise de agrupamento de Tocher.

O índice de seleção é considerado ferramenta indispensável no melhoramento genético quando se deseja obter ganhos simultâneos com a seleção. Essa seleção é realizada com base em diversas variáveis, o que resulta em maiores ganhos simultâneos (Freitas et al. 2012). O índice de seleção baseado na média de ranks (Mulamba and Mock 1978; Resende 2007) foi utilizado para auxiliar na escolha dos materiais superiores.

Os valores da Tabela 18 representam a posição (rank) que as 20 famílias híbridas assumiram para cada variável analisada e os ganhos com a seleção simultânea. Considerando uma intensidade de seleção de 50% (10 famílias) espera-se obter 16,99% de ganhos. Para um programa de melhoramento o ganho de 16,99% pode ser considerado satisfatório e de grande importância.

Foi realizada também uma análise de agrupamento multivariada das famílias, para todas as variáveis analisadas simultaneamente, pelo método de Tocher, baseado na distância genética euclidiana média. A análise de agrupamento separou as famílias em cinco grupos, sendo que no grupo 1 foram alocadas 14 das 20 famílias avaliadas (Tabela 19).

Com base nas análises de índice de seleção e agrupamento genético, as famílias H097, H0912, H0918 e H0921 se mostraram promissoras por estarem entre as 4 famílias com maiores ganhos simultâneos (Tabela 18) e terem sido alocadas em grupos distintos, o que indica maior variabilidade genética. Dessa forma, estas famílias podem ser inter cruzadas para avanço da seleção recorrente. Além disso, as famílias selecionadas podem ser avaliadas em diferentes regiões e, as que se destacarem, serem utilizadas como variedades de propagação por sementes. Para se obter novamente essas famílias serão selecionadas os genitores que deram origem as mesmas e estes serão inter cruzados, de modo que se espera que a média das famílias selecionadas se mantenha. Com essa estratégia espera-se evitar problemas de incompatibilidade genética e desenvolvimento de variedades mais tolerantes a seca que variedades clonais, associadas à característica de resistência à ferrugem, uniformidade de maturação e maior tamanho de fruto.

Análises conjunta de índice de seleção

Considerando as variáveis que apresentaram efeitos genéticos significativos para as famílias, de acordo com a análise de deviance (Tabela 15), foi realizada uma análise conjunta de índice de seleção de Mulamba-Mock para as famílias híbridas e seus respectivos genitores, Conilon e Robusta. Os valores da Tabela 20 representam a posição (rank) que as famílias e seus genitores assumiram para cada variável analisada e os ganhos com a seleção simultânea.

Com uma intensidade de seleção em torno de 10% (3 genótipos) espera-se obter 51,26% de ganhos. Por meio dessa análise foi possível verificar que as famílias H0912, H0918 e H0921 que se destacaram por proporcionarem os maiores ganhos com a seleção na análise das famílias também apresentaram maiores ganhos que os genitores que os originaram, demonstrando a ocorrência de heterose e heterobeltiose. Esse resultado evidencia a importância dos grupos heteróticos na obtenção de híbridos superiores, explorando os benefícios proporcionados pelos efeitos dos desvios de dominância. Dessa forma, essas famílias devem ser as selecionadas para continuar o programa de melhoramento tanto para obtenção de variedade propagadas por semente quanto para variedades clonais.

Seleção de Híbridos para as variáveis incidência à ferrugem e produção por planta

Na tabela 21 está apresentado os 20 melhores Híbridos dentro das famílias híbridas, para a variável incidência à ferrugem do cafeeiro. A incidência de ferrugem foi avaliada com notas de 1 a 5. A nota 1 foi atribuída para genótipos sem nenhum sintoma do patógeno e nota 5 para genótipos altamente suscetíveis. Dessa forma, quanto menor a nota atribuída menor a incidência de ferrugem e quanto menor o valor da nova média em relação à média geral (Tabela 16) mais resistente o genótipo.

Os cinco melhores híbridos pertencem à mesma família (H098). A seleção desses cinco genótipos permitirá um ganho genético de 0,16 sobre a média geral e, portanto, a nova média será de apenas 1,71. O valor genético aditivo é 1,25 e quanto menor esse valor melhor a seleção para essa variável. Dessa forma esses cinco Híbridos são de interesse para os programas de melhoramento genético que visam obter genótipos com menores níveis de incidência de ferrugem.

A seleção dos 20 melhores Híbridos permitirá um ganho genético de 0,19. A média da população melhorada em uma próxima geração de plantio será de 1,74.

Na tabela 22 está apresentado os 20 melhores Híbridos dentro das famílias híbridas, para a variável produção por planta. O Híbrido H0910-3 se destaca como o melhor genótipo dentre os

demais para a variável produção por planta. O Híbrido H0910-3 pertence à família H0910 e seu valor fenotípico observado (f) é de 21,00 litros por planta e o valor genético aditivo é de 10,11. Dessa forma, em uma reprodução via sementes desse genótipo, metade desse valor genético aditivo será transmitido para a sua descendência.

A seleção dos 20 melhores Híbridos permitirá um ganho genético de 1,49 litros por plantas sobre a média geral de 8,24 (Tabela 16). A média da população melhorada em uma próxima geração de plantio será de 9,74 litros por planta permitindo um ganho genético expressivo de 18,20%.

Os genótipos Híbridos que se destacam pelos menores níveis de incidência de ferrugem e maiores produções por planta podem ser clonados e utilizados nos programas de melhoramento genético do cafeeiro que buscam obter ganhos para essas variáveis.

Conclusões

Os genótipos Conilon UFV 3628-45, UFV 3627-27, UFV 3628-22 e UFV 3629-34, bem como os genótipos Robusta UFV 3360-169, UFV 3630-5 e UFV 3368-58 são promissores por serem geneticamente divergentes pela análise de agrupamento e por proporcionarem ganhos elevados com a seleção, podendo serem inter cruzados para obter cultivares dentro de cada grupo varietal ou como genitores em cruzamentos interpopulacionais.

As famílias híbridas H097, H0912, H0918 e H0921 que se destacaram nas análises podem ser testadas em diferentes regiões e constituírem variedade propagada por semente. Essas famílias têm potencial também para ser inter cruzadas (cruzamento intrapopulacional) para avanço da seleção recorrente.

Os Híbridos selecionados com menores incidência de ferrugem e com maiores produções por planta podem ser clonados e utilizados nos programas de melhoramento genético do cafeeiro.

Tabela 1. Genótipos do grupo varietal Conilon

Genótipos					
UFV 3627-8	UFV 3628-5	UFV 3628-32	UFV 3628-46	UFV 3629-12	UFV 3629-30
UFV 3627-20	UFV 3628-10	UFV 3628-33	UFV 3628-47	UFV 3629-15	UFV 3629-31
UFV 3627-24	UFV 3628-16	UFV 3628-35	UFV 3628-48	UFV 3629-16	UFV 3629-32
UFV 3627-25	UFV 3628-17	UFV 3628-36	UFV 3628-51	UFV 3629-17	UFV 3629-34
UFV 3627-27	UFV 3628-20	UFV 3628-37	UFV 3628-52	UFV 3629-20	UFV 3629-36
UFV 3627-29	UFV 3628-22	UFV 3628-38	UFV 3629-1	UFV 3629-23	UFV 3629-37
UFV 3627-30	UFV 3628-23	UFV 3628-39	UFV 3629-4	UFV 3629-24	UFV 3629-38
UFV 3627-31	UFV 3628-24	UFV 3628-40	UFV 3629-7	UFV 3629-25	UFV 3629-39
UFV 3628-1	UFV 3628-26	UFV 3628-42	UFV 3629-8	UFV 3629-26	UFV 3629-?
UFV 3628-2	UFV 3628-27	UFV 3628-43	UFV 3629-9	UFV 3629-27	UFV 3638-49
UFV 3628-3	UFV 3628-28	UFV 3628-44	UFV 3629-10	UFV 3629-28	UFV 513
UFV 3628-4	UFV 3628-29	UFV 3628-45	UFV 3629-11	UFV 3629-29	--

Tabela 2. Genótipos do grupo varietal Robusta

Genótipos				
UFV 3356-71	UFV 3363-125	UFV 3370-47	UFV 3376-9	UFV Apoatã-1
UFV 3356-74	UFV 3365-144	UFV 3370-49	UFV 3377-12	UFV Apoatã-2
UFV 3356-76	UFV 3366-134	UFV 3370-50	UFV 3630-10	UFV Apoatã-3
UFV 3357-91	UFV 3366-138	UFV 3371-19	UFV 3630-11	UFV Apoatã-4
UFV 3357-93	UFV 3366-139	UFV 3371-20	UFV 3630-2	UFV Apoatã-5
UFV 3358-88	UFV 3367-101	UFV 3371-22	UFV 3630-5	UFV Apoatã-6
UFV 3360-169	UFV 3367-105	UFV 3373-36	UFV 3630-6	UFV Apoatã-8
UFV 3360-171	UFV 3367-96	UFV 3373-43	UFV 3630-7	UFV 514
UFV 3361-148	UFV 3367-97	UFV 3374-29	UFV 3631-10	
UFV 3361-151	UFV 3367-98	UFV 3375-65	UFV 3631-11	
UFV 3362-118	UFV 3368-52	UFV 3375-66	UFV 3631-13	
UFV 3363-122	UFV 3368-58	UFV 3376-8	UFV 3631-9	

Tabela 3. Famílias híbridas e a descrição dos cruzamentos para obtenção das mesmas

Família	Cruzamentos
H092	UFV 3366-139 × UFV 513
H093	UFV 3366-139 × UFV 3629-11
H094	UFV 3367-98 × UFV 513
H095	UFV 3366-139 × UFV 3627-31
H097	UFV 3374-28 × UFV 3627-31
H098	UFV 3365-144 × UFV 3629-11
H099	UFV 3365-144 × UFV 513
H0910	UFV 3367-98 × UFV 3629-11
H0911	UFV 3367-98 × UFV 3628-2
H0912	UFV 3366-139 × UFV 3628-2
H0913	UFV 3365-144 × UFV 3628-2
H0914	UFV 3374-28 × UFV 513
H0915	UFV 3373-36 × UFV 3628-2
H0916	UFV 3367-98 × UFV 3629-25
H0917	UFV 3365-144 × UFV 3629-25
H0918	UFV 3373-36 × UFV 513
H0919	UFV 3373-36 × UFV 3629-11
H0920	UFV 3373-36 × UFV 3629-25
H0921	UFV 3374-28 × UFV 3629-11
H0922	UFV 3373-36 × UFV 3627-31

Tabela 4. Análise de Deviance (ANADEV) e teste da razão de verossimilhança (TRV) para as variáveis avaliadas na população de Conilon

Variáveis	Deviance				TRV (X ²)		
	Modelo	Genotípico	Blocos	Permanente	Genotípico	Blocos	Permanente
Vig	2878,78	2894,50	2938,41	3002,77	15,72**	59,63**	123,99**
Fer	1002,46	1065,39	1014,77	1014,70	62,93**	12,31**	12,24**
Cer	-1799,4	-1762,85	-1794,98	-1799,26	36,55**	4,42*	0,14 ^{NS}
AP	13208,80	13214,66	13267,39	13563,83	5,86*	58,59**	355,03**
DC	14101,69	14106,77	14174,04	14258,11	5,08*	72,35**	156,42**
Mat	516,76	587,29	519,57	517,36	70,53**	2,81 ^{NS}	0,60 ^{NS}
TF	-367,56	-268,33	-359,93	-365,18	99,23**	7,63**	2,38 ^{NS}
Prod	5288,05	5300,48	5298,06	5295,79	12,43**	10,01**	7,74**

** e * significativo a 1% e 5% de probabilidade, respectivamente, e ^{NS} não-significativo com base na tabela de Qui-quadrado com 1 grau de liberdade. Vig: vigor vegetativo; Fer: incidência à ferrugem do cafeeiro; Cer: incidência de cercosporiose; AP: altura da planta; DC: diâmetro da projeção da copa; Mat: época de maturação dos frutos; TF: tamanho de fruto; Prod: produção por planta. Genotípico: efeitos genotípicos; Blocos: efeitos de bloco; Permanente: efeitos de ambiente permanente.

Tabela 5. Estimativa dos parâmetros genéticos e ambientais para as variáveis avaliadas na população de Conilon

Parâmetros	Variáveis							
	Genéticos	Vig	Fer	Cer	AP	DC	Mat	TF
V_g	0,14	0,10	0,01	26,45	21,67	0,10	0,08	1,53
V_{bloc}	0,20	0,01	0,00	78,75	95,63	0,00	0,00	0,56
V_{perm}	0,40	0,04	0,00	225,45	164,34	0,01	0,01	1,77
V_e	1,32	0,54	0,11	269,27	485,08	0,49	0,23	22,40
V_f	2,06	0,69	0,13	599,92	766,71	0,60	0,33	26,26
h^2_g	0,07	0,14	0,08	0,04	0,03	0,16	0,25	0,06
r	0,36	0,21	0,09	0,55	0,37	0,19	0,29	0,15
c^2_{bloc}	0,10	0,02	0,01	0,13	0,12	0,01	0,01	0,02
c^2_{perm}	0,19	0,06	0,01	0,38	0,21	0,02	0,03	0,07
MG	5,79	2,06	2,08	146,64	127,94	1,91	1,98	8,11
$CV_{gi}(\%)$	6,52	15,14	4,70	3,51	3,64	16,27	14,43	15,26
$CV_e(\%)$	19,86	35,79	16,31	11,19	17,21	36,65	24,33	58,32
CV_{relativo}	0,33	0,42	0,29	0,31	0,21	0,44	0,59	0,26
r_{gg}	0,70	0,86	0,78	0,56	0,53	0,84	0,89	0,67

V_g : Variância genética; V_{bloc} : Variância ambiental entre blocos; V_{perm} : Variância ambiente permanente; V_e : Variância residual; V_f : Variância fenotípica; h^2_g : Herdabilidade individual no sentido amplo (efeitos genotípicos totais); r : Repetibilidade; c^2_{bloc} : Coeficiente de determinação ambiental entre blocos; c^2_{perm} : Coeficiente de determinação de ambiente permanente; MG: Média Geral; $CV_{gi}(\%)$: Coeficiente de variação genotípico; $CV_e(\%)$: Coeficiente de variação residual; CV_{relativo} : $CV_{gi}(\%)/CV_e(\%)$; r_{gg} : Acurácia seletiva. Vig: vigor vegetativo; Fer: incidência à ferrugem do cafeeiro; Cer: incidência de cercosporiose; AP: altura da planta; DC: diâmetro da projeção da copa; Mat: época de maturação dos frutos; TF: tamanho de fruto; Prod: produção por planta.

Tabela 6. Correlação genética entre as variáveis avaliadas na população de Conilon

	Vig	Fer	Cer	AP	DC	Mat	TF	Prod
Vig	1,00	-0,38	-0,55	0,58	0,69	-0,04	0,05	0,42
Fer		1,00	0,57	-0,2	-0,26	-0,32	0,17	-0,2
Cer			1,00	-0,2	-0,29	-0,12	0,09	-0,57
AP				1,00	0,68	-0,23	0,24	0,18
DC					1,00	-0,16	0,15	0,38
Mat						1,00	-0,28	0,05
TF							1,00	0,28
Prod								1,00

Vig: vigor vegetativo; Fer: incidência à ferrugem do cafeeiro; Cer: incidência de cercosporiose; AP: altura da planta; DC: diâmetro da projeção da copa; Mat: época de maturação dos frutos; TF: tamanho de fruto; Prod: produção por planta.

Tabela 7. Classificação dos 15 genótipos Conilon superiores com base no *rank* médio (índice de Mulamba-Mock), referente a oito variáveis fenotípicas

Ordem	Conilon	Vig	Fer	Cer	AP	DC	Mat	TF	Prod	Rank-Médio	Ganho (%)
1	UFV 3628-2	3	3	3	71	1	2	32	7	15,25	136,07
2	UFV 3628-45	1	1	4	69	4	4	58	16	19,63	106,45
3	UFV 3627-27	9	14	5	43	17	18	41	13	20,00	96,81
4	UFV 3627-29	5	6	1	50	10	7	71	15	20,63	90,73
5	UFV 3629-26	24	22	12	42	15	47	13	6	22,63	83,44
6	UFV 3629-4	13	5	13	21	29	63	16	22	22,75	78,70
7	UFV 3628-22	8	52	39	66	7	5	2	12	23,88	74,09
8	UFV 3628-4	37	12	6	67	11	17	19	25	24,25	70,41
9	UFV 3629-34	2	9	23	64	3	54	37	5	24,63	67,33
10	UFV 3628-26	11	37	33	38	28	27	6	19	24,88	64,76
11	UFV 3628-28	38	59	49	11	2	12	3	26	25,00	62,63
12	UFV 3629-29	27	18	21	47	13	59	10	14	26,13	60,22
13	UFV 3629-27	33	15	11	24	35	65	23	4	26,25	58,17
14	UFV 3628-40	18	30	8	26	41	62	8	18	26,38	56,40
15	UFV 3628-46	6	32	16	23	9	64	35	27	26,50	54,84

Vig: vigor vegetativo; Fer: incidência à ferrugem do cafeeiro; Cer: incidência de cercosporiose; AP: altura da planta; DC: diâmetro da projeção da copa; Mat: época de maturação dos frutos; TF: tamanho de fruto; Prod: produção por planta.

Tabela 8. Agrupamento dos 71 genótipos Conilon pelo método de Tocher, com base na distância genética euclidiana média

Grupo	Conilon				
1	UFV 3629-4	UFV 3629-23	UFV 3628-46	UFV 3628-32	UFV 3629-30
	UFV 3627-27	UFV 3628-20	UFV 3627-20	UFV 3628-35	UFV 3629-10
	UFV 3628-23	UFV 3629-17	UFV 3629-27	UFV 3628-37	UFV 3628-16
	UFV 3629-?	UFV 3627-30	UFV 3628-42	UFV 3629-15	UFV 3629-32
	UFV 3628-5	UFV 3628-38	UFV 3628-44	UFV 3629-8	UFV 3628-39
	UFV 3629-11	UFV 513	UFV 3629-1	UFV 3627-25	UFV 3628-48
	UFV 3628-26	UFV 3628-3	UFV 3628-28	UFV 3629-25	UFV 3629-28
	UFV 3629-38	UFV 3629-36	UFV 3629-39	UFV 3628-40	UFV 3628-45
	UFV 3629-34	UFV 3629-29	UFV 3629-7	UFV 3638-49	UFV 3628-52
	UFV 3628-4	UFV 3628-24	UFV 3629-37	UFV 3629-9	UFV 3629-16
	UFV 3627-31	UFV 3629-12	UFV 3628-10	UFV 3629-31	UFV 3628-36
	UFV 3627-29	UFV 3629-24	UFV 3628-33	UFV 3629-20	UFV 3628-43
	UFV 3628-1	UFV 3628-47	UFV 3628-22	UFV 3627-24	UFV 3628-27
	UFV 3627-8	UFV 3628-17	UFV 3628-29	UFV 3629-26	UFV 3628-51
2	UFV 3628-2				

Tabela 9. Agrupamento dos 70 genótipos Conilon pelo método de Tocher, com base na distância genética euclidiana média

Grupo	Conilon			
1	UFV 3628-32	UFV 3628-35	UFV 3628-23	UFV 3628-37
	UFV 3629-15	UFV 3629-32	UFV 3628-44	UFV 3629-8
	UFV 3628-39	UFV 3628-48	UFV 3629-25	UFV 3629-38
	UFV 3629-36	UFV 3628-52	UFV 3628-24	UFV 3629-37
	UFV 3629-9	UFV 3627-31	UFV 3628-36	UFV 3628-33
	UFV 3628-43	UFV 3628-47	UFV 3628-27	UFV 3628-51
2	UFV 3629-4	UFV 3628-46	UFV 3629-30	UFV 3627-27
	UFV 3629-27	UFV 3627-30	UFV 3628-5	UFV 3628-26
	UFV 3628-40	UFV 3629-29	UFV 3628-4	UFV 3629-24
	UFV 3628-17	UFV 3629-26		
3	UFV 3629-23	UFV 3629-1	UFV 3629-28	UFV 3629-12
4	UFV 3628-16	UFV 3629-?	UFV 3628-42	UFV 513
	UFV 3628-3	UFV 3627-8		
5	UFV 3628-38	UFV 3638-49	UFV 3629-16	
6	UFV 3627-20	UFV 3629-10	UFV 3629-17	UFV 3627-25
7	UFV 3628-10	UFV 3628-22	UFV 3628-29	
8	UFV 3628-45	UFV 3627-29	UFV 3627-24	
9	UFV 3629-39	UFV 3629-7		
10	UFV 3628-28	UFV 3628-1		
11	UFV 3628-20			
12	UFV 3629-20			
13	UFV 3629-34			
14	UFV 3629-11			
15	UFV 3629-31			

Tabela 10. Análise de Deviance (ANADEV) e teste da razão de verossimilhança (TRV) para as variáveis avaliadas na população de Robusta

Variáveis	Deviance				TRV (X ²)		
	Modelo Completo	Genotípico	Blocos	Permanente	Genotípico	Blocos	Permanente
Vig	1787,14	1792,16	1788,65	1856,04	5,02*	1,51 ^{NS}	68,9**
Fer	-1210,9	-1180,72	-1210,88	-1210,93	30,18**	0,02 ^{NS}	-0,03 ^{NS}
Cer	-579,57	-558,39	-578,6	-574,85	21,18**	0,97 ^{NS}	4,72*
AP	8605	8610,05	8607,32	8723,2	5,05*	2,32 ^{NS}	118,2**
DC	8502,88	8506,16	8507,01	8612,65	3,28 ^{NS}	4,13*	109,77**
Mat	103,74	144,45	111,88	103,83	40,71**	8,14**	0,09 ^{NS}
TF	-369,05	-288,77	-369,03	-369,06	80,28**	0,02 ^{NS}	-0,01 ^{NS}
Prod	2920,41	2926,86	2922,27	2922,38	6,45*	1,86 ^{NS}	1,97 ^{NS}

** e * significativo a 1% e 5% de probabilidade, respectivamente, e ^{NS} não-significativo com base na tabela de Qui-quadrado com 1 grau de liberdade. Vig: vigor vegetativo; Fer: incidência à ferrugem do cafeeiro; Cer: incidência de cercosporiose; AP: altura da planta; DC: diâmetro da projeção da copa; Mat: época de maturação dos frutos; TF: tamanho de fruto; Prod: produção por planta. Genotípico: efeitos genotípicos; Blocos: efeitos de bloco; Permanente: efeitos de ambiente permanente.

Tabela 11. Estimativa dos parâmetros genéticos e ambientais para as variáveis avaliadas na população de Robusta

Parâmetros Genéticos	Variáveis							
	Vig	Fer	Cer	AP	DC	Mat	TF	Prod
V_g	0,11	0,01	0,02	73,88	45,37	0,09	0,07	2,32
V_{bloc}	0,03	0,00	0,00	21,60	30,92	0,03	0,00	0,70
V_{perm}	0,45	0,00	0,01	347,67	286,71	0,01	0,00	2,20
V_e	1,46	0,11	0,18	657,64	611,61	0,36	0,17	33,14
V_f	2,05	0,12	0,21	1100,78	974,61	0,48	0,24	38,36
h^2_g	0,05	0,07	0,09	0,07	0,05	0,19	0,29	0,06
r	0,29	0,08	0,16	0,40	0,37	0,26	0,30	0,14
c^2_{bloc}	0,01	0,00	0,01	0,02	0,03	0,05	0,00	0,02
c^2_{perm}	0,22	0,00	0,06	0,32	0,29	0,01	0,01	0,06
MG	5,75	1,21	2,07	158,46	135,88	2,55	2,52	7,07
CV_{gi} (%)	5,72	7,76	6,79	5,42	4,96	11,92	10,52	21,56
CV_e (%)	21,02	27,85	20,22	16,18	18,20	23,59	16,27	81,40
CV_{relativo}	0,27	0,28	0,34	0,34	0,27	0,51	0,65	0,26
r_{gg}	0,58	0,75	0,75	0,59	0,53	0,82	0,87	0,59

V_g : Variância genética; V_{bloc} : Variância ambiental entre blocos; V_{perm} : Variância ambiente permanente; V_e : Variância residual; V_f : Variância fenotípica; h^2_g : Herdabilidade individual no sentido amplo (efeitos genotípicos totais); r : Repetibilidade; c^2_{bloc} : Coeficiente de determinação ambiental entre blocos; c^2_{perm} : Coeficiente de determinação de ambiente permanente; MG: Média Geral; CV_{gi} (%): Coeficiente de variação genotípico; CV_e (%): Coeficiente de variação residual; CV_{relativo} : $CV_{gi}(\%)/CV_e(\%)$; r_{gg} : Acurácia seletiva. Vig: vigor vegetativo; Fer: incidência à ferrugem do cafeeiro; Cer: incidência de cercosporiose; AP: altura da planta; DC: diâmetro da projeção da copa; Mat: época de maturação dos frutos; TF: tamanho de fruto; Prod: produção por planta.

Tabela 12. Correlação genética entre as variáveis avaliadas na população de Robusta

	Vig	Fer	Cer	AP	DC	Mat	TF	Prod
Vig	1,00	0,02	-0,11	0,63	0,80	-0,16	0,01	0,36
Fer		1,00	-0,05	-0,08	0,04	-0,20	-0,20	0,24
Cer			1,00	-0,07	-0,11	-0,06	0,20	0,01
AP				1,00	0,68	-0,19	0,04	0,14
DC					1,00	-0,25	0,06	0,37
Mat						1,00	0,19	-0,10
TF							1,00	0,00
Prod								1,00

Vig: vigor vegetativo; Fer: incidência à ferrugem do cafeeiro; Cer: incidência de cercosporiose; AP: altura da planta; DC: diâmetro da projeção da copa; Mat: época de maturação dos frutos; TF: tamanho de fruto; Prod: produção por planta.

Tabela 13. Classificação dos 15 melhores genótipos Robusta com base no *rank* médio (índice de Mulamba-Mock), referente a sete variáveis fenotípicas

Ordem	Robusta	Vig	Fer	Cer	AP	Mat	TF	Prod	Rank-Médio	Ganho (%)
1	UFV 3631-10	18	8	17	7	13	47	13	17,57	62,20
2	UFV 3371-20	6	29	28	40	6	8	6	17,57	62,20
3	UFV Aboatã-2	1	1	4	54	32	2	32	18,00	60,89
4	UFV 3370-50	25	2	13	31	46	7	5	18,43	59,28
5	UFV 3357-91	44	4	9	12	34	35	3	20,14	55,37
6	UFV 3630-5	2	3	1	49	43	41	2	20,14	52,87
7	UFV 3630-6	7	28	16	41	24	3	31	21,43	49,68
8	UFV 3356-74	17	5	3	21	40	48	17	21,57	47,23
9	UFV 3368-52	23	22	34	9	26	37	9	22,86	44,33
10	UFV 3365-144	42	10	45	4	28	11	20	22,86	42,09
11	UFV Aboatã-1	15	6	7	53	20	30	35	23,71	39,78
12	UFV 3360-169	10	56	5	43	3	50	4	24,43	37,51
13	UFV 3368-58	11	51	56	44	5	6	1	24,86	35,43
14	UFV 3373-36	9	12	35	56	4	43	16	25,00	33,64
15	UFV 3358-88	26	31	24	2	36	29	30	25,43	31,94

Vig: vigor vegetativo; Fer: incidência à ferrugem do cafeeiro; Cer: incidência de cercosporiose; AP: altura da planta; Mat: época de maturação dos frutos; TF: tamanho de fruto; Prod: produção por planta.

Tabela 14. Agrupamento dos 56 genótipos Robusta pelo método de Tocher, com base na distância genética euclidiana média

Grupo	Robusta				
1	UFV 3366-134	UFV 3366-139	UFV 3357-91	UFV 3376-8	UFV 3371-22
	UFV 3375-65	UFV 3368-52	UFV 3371-19	UFV Apoatã-1	UFV 3370-47
	UFV 3631-13	UFV 514	UFV 3373-36	UFV 3631-9	UFV Apoatã-3
	UFV Apoatã-8	UFV 3356-76	UFV 3367-105	UFV 3356-71	UFV 3370-50
	UFV 3361-148	UFV 3367-98	UFV 3363-122	UFV Apoatã-5	UFV 3631-11
	UFV 3630-11	UFV 3367-101	UFV 3630-6	UFV 3376-9	UFV 3367-97
	UFV 3366-138	UFV Apoatã-2	UFV 3375-66	UFV 3631-10	UFV 3357-93
	UFV 3377-12	UFV 3373-43	UFV 3370-49	UFV 3374-29	UFV Apoatã-4
	UFV 3363-125	UFV Apoatã-6	UFV 3360-171	UFV 3630-10	UFV 3371-20
	UFV 3356-74	UFV 3358-88	UFV 3362-118	UFV 3365-144	UFV 3361-151
	UFV 3367-96	UFV 3630-2			
2	UFV 3360-169	UFV 3630-7			
3	UFV 3630-5				
4	UFV 3368-58				

Tabela 15. Análise de Deviance (ANADEV) e teste da razão de verossimilhança (TRV) para as variáveis avaliadas nas famílias híbridas

Variáveis	Deviance						TRV (X ²)				
	Modelo Completo	CEC	Fêmeas	Machos	Permanente	Blocos	CEC	Fêmeas	Machos	Permanente	Blocos
Vig	1272,16	1278,42	1275,34	1272,15	1444,70	1274,78	6,26*	3,18 ^{NS}	-0,01 ^{NS}	172,54**	2,62 ^{NS}
Fer	765,49	766,02	772,23	766,01	819,10	765,62	0,53 ^{NS}	6,74**	0,52 ^{NS}	53,61**	0,13 ^{NS}
AP	8607,12	8615,25	8607,12	8608,74	8748,84	8610,42	8,13**	0,00 ^{NS}	1,62 ^{NS}	141,72**	3,30 ^{NS}
DC	9122,62	9125,39	9123,06	9122,61	9249,89	9123,65	2,77 ^{NS}	0,44 ^{NS}	-0,01 ^{NS}	127,27**	1,03 ^{NS}
Mat	557,88	558,56	558,14	564,48	564,20	557,96	0,68 ^{NS}	0,26 ^{NS}	6,60*	6,32*	0,08 ^{NS}
TF	-336,83	-336,83	-336,01	-336,83	-336,82	-333,43	0,00 ^{NS}	0,82 ^{NS}	0,00 ^{NS}	0,01 ^{NS}	3,40 ^{NS}
Prod	3480,63	3483,77	3480,63	3480,63	3480,67	3485,53	3,14 ^{NS}	0,00 ^{NS}	0,00 ^{NS}	0,04 ^{NS}	4,90*

** e * significativo a 1% e 5% de probabilidade, respectivamente, e ^{NS} não-significativo com base na tabela de Qui-quadrado com 1 grau de liberdade. Vig: vigor vegetativo; Fer: incidência à ferrugem do cafeeiro; AP: altura da planta; DC: diâmetro da projeção da copa; Mat: época de maturação dos frutos; TF: tamanho de fruto; Prod: produção por planta. CEC: efeitos da capacidade específica de combinação dos genitores da população de fêmeas com os genitores da população de machos; Fêmeas: efeitos de genitores da população de fêmeas; Machos: efeitos de genitores da população de machos; Permanente: efeitos permanentes de indivíduo; Blocos: efeitos de bloco.

Tabela 16. Estimativa dos parâmetros genéticos e ambientais para as variáveis avaliadas nas famílias híbridas

Parâmetros	Variáveis						
	Vig	Fer	AP	DC	Mat	TF	Prod
V_{gm}	0,00	0,02	69,72	1,44	0,08	0,00	0,01
V_{gf}	0,12	0,07	1,17	19,85	0,01	0,00	0,01
V_{fm}	0,09	0,01	67,49	45,08	0,01	0,00	0,90
V_{perm}	0,47	0,14	227,91	325,55	0,07	0,00	0,30
V_{bloc}	0,03	0,00	28,05	17,37	0,00	0,01	1,13
V_e	0,80	0,58	466,39	753,13	0,67	0,15	35,97
V_f	1,51	0,83	860,73	1162,43	0,85	0,15	38,32
c^2_m	0,00	0,02	0,08	0,00	0,10	0,00	0,00
c^2_f	0,08	0,09	0,00	0,02	0,01	0,01	0,00
c^2_{fm}	0,06	0,01	0,08	0,04	0,02	0,00	0,02
c^2_{perm}	0,31	0,17	0,26	0,28	0,08	0,01	0,01
c^2_{bloc}	0,02	0,01	0,03	0,01	0,01	0,04	0,03
r	0,47	0,29	0,46	0,35	0,21	0,05	0,06
h^2_{am}	0,00	0,08	0,32	0,00	0,38	0,00	0,00
h^2_{af}	0,32	0,35	0,01	0,07	0,03	0,03	0,00
h^2_a	0,16	0,21	0,16	0,04	0,21	0,02	0,00
h^2_{dom}	0,23	0,06	0,31	0,16	0,07	0,00	0,09
h^2_g	0,39	0,27	0,48	0,19	0,27	0,02	0,10
MG	6,26	1,45	174,26	158,24	2,02	2,13	8,24
CV_{gi} (%)	4,67	7,40	4,71	4,24	5,84	0,25	11,53
CV_e (%)	14,30	52,72	12,39	17,34	40,47	17,91	72,78
$CV_{relativo}$	0,33	0,14	0,38	0,24	0,14	0,01	0,16
r_{gg}	0,60	0,39	0,67	0,57	0,42	0,05	0,55

V_{gm} : variância genética entre genitores da população de machos em cruzamento com a população de fêmeas, a qual estima (1/4) da variação genética aditiva correspondente; V_{gf} : variância genética entre genitores da população de fêmeas em cruzamento com a população machos, a qual estima (1/4) da variação genética aditiva correspondente; V_{fm} : variância da capacidade específica de combinação interpopulacional entre dois genitores, a qual estima (1/4) da variação genética de dominância correspondente; V_{perm} : variância dos efeitos permanentes; V_{bloc} : variância entre blocos; V_e : variância residual; V_f : variância fenotípica individual; c^2_m : coeficiente de determinação dos efeitos de genitores da população de machos; c^2_f : coeficiente de determinação dos efeitos de genitores da população de fêmeas; c^2_{fm} : coeficiente de determinação dos efeitos da capacidade específica de combinação; c^2_{perm} : coeficiente de determinação dos efeitos permanentes; c^2_{bloc} : coeficiente de determinação dos efeitos de bloco; r : repetibilidade individual; h^2_{am} : herdabilidade individual no sentido restrito na população de machos, ou seja, dos efeitos aditivos interpopulacionais; h^2_{af} : herdabilidade individual no sentido restrito na população de fêmeas, ou seja, dos efeitos aditivos interpopulacionais; h^2_{dom} : herdabilidade individual dos efeitos interpopulacionais de dominância; h^2_a : herdabilidade interpopulacional individual no sentido restrito, média para as duas populações; h^2_g : herdabilidade interpopulacional individual no sentido amplo, ou seja, dos efeitos genotípicos totais; MG: Média Geral do experimento; CV_{gi} (%): Coeficiente de variação genotípico; CV_e (%): Coeficiente de variação residual; $CV_{relativo}$: $CV_{gi}(\%)/CV_e(\%)$; r_{gg} : Acurácia seletiva. Vig: vigor vegetativo; Fer: incidência à ferrugem do cafeeiro; AP: altura da planta; DC: diâmetro da projeção da copa; Mat: época de maturação dos frutos; TF: tamanho de fruto; Prod: produção por planta.

Tabela 17. Correlação genética entre as variáveis avaliadas nas famílias híbridas

	Vig	Fer	AP	DC	Mat	TF	Prod
Vig	1,00	0,23	0,82	0,88	-0,15	0,07	0,32
Fer		1,00	0,29	0,34	-0,50	0,26	0,23
AP			1,00	0,86	-0,17	0,29	0,30
DC				1,00	-0,21	0,12	0,44
Mat					1,00	-0,31	-0,01
TF						1,00	-0,13
Prod							1,00

Vig: vigor vegetativo; Fer: incidência à ferrugem do cafeeiro; AP: altura da planta; DC: diâmetro da projeção da copa; Mat: época de maturação dos frutos; TF: tamanho de fruto; Prod: produção por planta.

Tabela 18. Classificação das famílias híbridas com base no *rank* médio (índice de Mulamba-Mock), referente a quatro variáveis fenotípicas

Ordem	Família	Vig	Fer	AP	Mat	Rank-Médio	Ganho (%)
1	H0921	2	8	17	3	7,50	40,00
2	H0912	9	1	8	13	7,75	37,70
3	H0918	12	7	9	5	8,25	34,04
4	H097	14	2	1	19	9,00	29,23
5	H0917	18	5	3	10	9,00	26,51
6	H099	4	4	12	17	9,25	24,14
7	H0919	16	6	2	14	9,50	21,99
8	H0913	15	16	7	1	9,75	20,00
9	H095	3	20	14	2	9,75	18,50
10	H0922	1	15	20	4	10,00	16,99
11	H0916	6	9	13	15	10,75	14,93
12	H0915	10	10	11	12	10,75	13,26
13	H092	20	12	5	7	11,00	11,66
14	H098	19	11	6	8	11,00	10,32
15	H0914	13	17	10	6	11,50	8,81
16	H0910	7	3	18	20	12,00	7,18
17	H0920	17	14	4	16	12,75	5,31
18	H093	5	18	19	9	12,75	3,70
19	H0911	8	13	15	18	13,50	1,92
20	H094	11	19	16	11	14,25	0,00

Vig: vigor vegetativo; Fer: incidência à ferrugem do cafeeiro; AP: altura da planta; Mat: época de maturação dos frutos.

Tabela 19. Agrupamento das 20 famílias híbridas pelo método de Tocher, com base na distância genética euclidiana média

Grupo	Famílias						
1	H094	H0916	H0920	H093	H0915	H0914	H0913
	H0918	H0911	H0917	H099	H092	H098	H0919
2	H0921	H0922					
3	H097	H0910					
4	H0912						
5	H095						

Tabela 20. Classificação das famílias híbridas e seus genitores com base no *rank* médio (índice de Mulamba-Mock), referente a quatro variáveis fenotípicas

Ordem	Genótipo/Família	Vig	Fer	AP	Mat	Rank-Médio	Ganho (%)
1	H0921	1	8	20	5	8,50	76,47
2	H0912	2	13	13	11	9,75	64,38
3	H0918	14	10	16	6	11,50	51,26
4	UFV 3365-144	16	1	15	19	12,75	41,18
5	UFV 3627-31	17	26	7	2	13,00	35,14
6	H0913	28	9	8	7	13,00	31,39
7	H0917	29	5	1	18	13,25	28,44
8	UFV 3373-36	9	2	29	15	13,75	25,65
9	H098	27	7	5	17	14,00	23,29
10	H0911	3	15	17	22	14,25	21,21
11	H0922	6	19	28	4	14,25	19,57
12	UFV 3629-25	21	21	6	10	14,50	18,03
13	H0915	13	11	19	16	14,75	16,59
14	H099	25	6	14	14	14,75	15,38
15	H0919	24	12	3	20	14,75	14,36
16	UFV 513	23	23	11	3	15,00	13,34
17	H0916	7	18	9	26	15,00	12,46
18	UFV 3366-139	12	4	18	27	15,25	11,57
19	H097	11	16	12	24	15,75	10,57
20	H0914	18	14	23	8	15,75	9,69
21	UFV 3367-98	10	3	26	29	17,00	8,43
22	UFV 3628-2	19	22	27	1	17,25	7,23
23	H092	22	28	10	9	17,25	6,15
24	H095	4	29	24	13	17,50	5,11
25	UFV 3629-11	20	24	4	23	17,75	4,09
26	H094	15	25	21	12	18,25	3,04
27	H0920	26	20	2	25	18,25	2,08
28	H0910	5	17	25	28	18,75	1,08
29	H093	8	27	22	21	19,50	0,00

Vig: vigor vegetativo; Fer: incidência à ferrugem do cafeeiro; AP: altura da planta; Mat: época de maturação dos frutos.

Tabela 21. Seleção dos 20 melhores Híbridos dentro das famílias híbridas, para a variável incidência à ferrugem, com base nos efeitos aditivos (a) e valores genéticos aditivos (u + a)

Ordem	Híbrido	Família	f	a	u+a	Ganho	Nova Média
1	H098-12	H098	1,00	-0,33	1,23	0,15	1,71
2	H098-14	H098	1,00	-0,31	1,24	0,16	1,71
3	H098-2	H098	1,00	-0,30	1,25	0,16	1,71
4	H098-3	H098	1,00	-0,30	1,25	0,16	1,71
5	H098-10	H098	1,00	-0,30	1,25	0,16	1,71
6	H099-8	H099	1,00	-0,30	1,25	0,16	1,72
7	H0917-2	H0917	1,00	-0,29	1,26	0,16	1,72
8	H0917-3	H0917	1,00	-0,29	1,26	0,17	1,72
9	H098-5	H098	1,00	-0,29	1,26	0,17	1,72
10	H099-14	H099	1,00	-0,29	1,26	0,17	1,72
11	H098-1	H098	1,00	-0,28	1,27	0,17	1,72
12	H099-9	H099	1,00	-0,28	1,27	0,17	1,73
13	H099-2	H099	1,00	-0,28	1,27	0,18	1,73
14	H098-11	H098	1,00	-0,28	1,28	0,18	1,73
15	H099-3	H099	1,00	-0,28	1,28	0,18	1,73
16	H099-10	H099	1,00	-0,27	1,28	0,18	1,73
17	H0917-1	H0917	1,00	-0,27	1,28	0,18	1,74
18	H099-5	H099	1,00	-0,26	1,29	0,19	1,74
19	H099-4	H099	1,00	-0,26	1,29	0,19	1,74
20	H099-1	H099	1,00	-0,26	1,29	0,19	1,74

f: valor fenotípico individual ou medição de campo; a: efeito genético aditivo predito; u + a: valor genético aditivo predito.

Tabela 22. Seleção dos 20 melhores Híbridos dentro das famílias híbridas, para a variável produção por planta, com base nos efeitos aditivos (a) e valores genéticos aditivos (u + a)

Ordem	Híbrido	Família	f	a	u+a	Ganho	Nova Média
1	H0910-3	H0910	21,00	1,86	10,11	1,86	10,11
2	H095-23	H095	14,88	1,70	9,94	1,78	10,03
3	H095-15	H095	13,63	1,66	9,90	1,74	9,98
4	H095-4	H095	11,50	1,55	9,80	1,69	9,94
5	H095-9	H095	14,33	1,54	9,79	1,66	9,91
6	H095-26	H095	12,38	1,54	9,78	1,64	9,89
7	H095-30	H095	14,00	1,51	9,75	1,62	9,87
8	H095-31	H095	11,75	1,50	9,75	1,61	9,85
9	H095-18	H095	24,00	1,49	9,74	1,60	9,84
10	H095-1	H095	11,00	1,49	9,74	1,58	9,83
11	H095-32	H095	11,38	1,47	9,71	1,57	9,82
12	H095-29	H095	13,00	1,44	9,68	1,56	9,81
13	H095-3	H095	9,13	1,43	9,67	1,55	9,80
14	H095-2	H095	10,63	1,42	9,66	1,54	9,79
15	H0952-8	H095	11,88	1,42	9,66	1,53	9,78
16	H095-11	H095	10,25	1,39	9,64	1,53	9,77
17	H095-21	H095	19,00	1,39	9,63	1,52	9,76
18	H0910-1	H0910	15,83	1,39	9,63	1,51	9,75
19	H095-7	H095	12,00	1,36	9,61	1,50	9,75
20	H095-8	H095	9,75	1,34	9,59	1,49	9,74

f: valor fenotípico individual ou medição de campo; a: efeito genético aditivo predito; u + a: valor genético aditivo predito.

Referências

- Babova O, Occhipinti A, Maffei ME (2016) Chemical partitioning and antioxidant capacity of green coffee (*Coffea arabica* and *Coffea canephora*) of different geographical origin. *Phytochemistry* 123:33–39. doi: 10.1016/j.phytochem.2016.01.016
- Bicho NC, Lidon FC, Ramalho JC, Leitão AE (2013) Quality assessment of Arabica and Robusta green and roasted coffees – A review. *Emirates J Food Agric* 25:945–950. doi: 10.9755/ejfa.v25i12.17290
- Carias CM de OM, Gravina G do A, Ferrão MAG, et al (2016) Predição de ganhos genéticos via modelos mistos em progênies de café Conilon. *Coffee Sci* 11:39–45.
- Charrier A, Berthaud J (1988) Principles and methods in coffee plant breeding: *Coffea canephora* Pierre. In: Clarke R, Macrae R (eds) *Coffee*. Elsevier Applied Science, London and New York, Great Britain, pp 167–198
- Costa RB Da, Resende MDV De, Gonçalves PDS, et al (2010) Predição de parâmetros e valores genéticos para caracteres de crescimento e produção de látex em progênies de seringueira. *Bragantia* 69:49–56. doi: <http://dx.doi.org/10.1590/S0006-87052010000100007>
- 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
- de Resende MDV, Ramalho MAP, Guilherme SR, de F.B. Abreu Â (2015) Multigeneration Index in the Within-Progenies Bulk Method for Breeding of Self-pollinated Plants. *Crop Sci* 55:1202–1211. doi: 10.2135/cropsci2014.08.0580
- Falconer DS, Mackay TFC (1996) *Introduction to Quantitative Genetics*, 4.ed. Edinburgh
- Ferrão L, Caixeta E, Souza F, et al (2013) Comparative study of different molecular markers for classifying and establishing genetic relationships in *Coffea canephora*. *Plant Syst Evol* 299:225–238. doi: 10.1007/s00606-012-0717-2
- Ferreira A, Cecon PR, Cruz CD, et al (2005) Seleção simultânea de *Coffea canephora* por meio da combinação de análise de fatores e índices de seleção. *Pesq agropec bras* 40:1189–1195.
- Fonseca AFA da, Sedyama T, Cruz CD, et al (2004) Repeatability and number of harvests required for selection in robusta coffee. *Crop Breed Appl Biotechnol* 4:325–329.
- Freitas JPX de, Oliveira EJ de, Jesus ON de, et al (2012) Formação de população base para seleção recorrente em maracujazeiro-amarelo com uso de índices de seleção. *Pesqui Agropecuária Bras* 47:393–401. doi: 10.1590/S0100-204X2012000300011

- Grandillo S (2014) Introgression Libraries with Wild Relatives of Crops. *Genomics Plant Genet Resour* 2:87–122. doi: 10.1007/978-94-007-7575-6_4
- Henderson CR (1975) Best linear unbiased estimation and prediction under a selection model. *Biometrics* 31:423–449.
- Leroy T, Marraccini P, Dufour M, et al (2005) Construction and characterization of a *Coffea canephora* BAC library to study the organization of sucrose biosynthesis genes. *Theor Appl Genet* 111:1032–1041. doi: 10.1007/s00122-005-0018-z
- Maia MCC, Resende MDV de, Oliveira LC de, et al (2011) Seleção de clones experimentais de cupuaçu para características agroindustriais via modelos mistos. *Rev AGRO@MBIENTE ON-LINE* 5:35–43. doi: <http://dx.doi.org/10.18227/1982-8470ragro.v5i1.388>
- Marraccini P, Vinecky F, Alves GSC, et al (2012) Differentially expressed genes and proteins upon drought acclimation in tolerant and sensitive genotypes of *Coffea canephora*. *J Exp Bot* 63:4191–4212. doi: 10.1093/jxb/ers103
- Mistro JC (2013) Estimativas de parâmetros genéticos visando o melhoramento do café robusta (*Coffea canephora* Pierre ex. A. Froehner). Escola Superior de Agricultura “Luiz de Queiroz”, Piracicaba
- Mrode RA (2014) *Linear Models for the Prediction of Animal Breeding Values*, 3rd edn.
- Mulamba NN, Mock JJ (1978) Improvement of yield potential of the Eto Blanco maize (*Zea mays* L.) population by breeding for plant traits. *J Genet Cytol* 7:40–51.
- Musoli P, Cubry P, Aluka P, et al (2009) Genetic differentiation of wild and cultivated populations: diversity of *Coffea canephora* Pierre in Uganda. *Genome* 52:634–646. doi: 10.1139/G09-037
- Oliveira EJ, Santana FA, Oliveira LA, Santos VS (2014) Genetic parameters and prediction of genotypic values for root quality traits in cassava using REML/BLUP. *Genet Mol Res* 13:6683–6700. doi: 10.4238/2014.August.28.13
- Oliveira EJ de, Santos V da S, Lima DS de, et al (2011) Estimativas de correlações genótípicas e fenótípicas em germoplasma de maracujazeiro. *Bragantia* 70:255–261.
- 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
- Pimentel-Gomes F (2009) *Curso de Estatística Experimental*. Piracicaba: FEALQ
- Pompeu Junior J, Blumer S, Resende MDV de (2013) Avaliação genética de seleções e híbridos de limões cravo, volkameriano e rugoso como porta-enxertos para laranjeiras Valência na

- presença da morte súbita dos citros. *Rev Bras Frutic* 35:199–209. doi: 10.1590/S0100-29452013000100023
- Ramalho AR, Rocha RB, Souza FF, et al (2016) Progresso genético da produtividade de café beneficiado com a seleção de clones de cafeeiro “Conilon.” *Rev Ciência Agronômica* 47:516–523. doi: 10.5935/1806-6690.20160062
- Rao RC (1952) *Advanced statistical methods in biometric research*. New York
- Resende RMS, Jank L, Valle CB do, Bonato ALV (2004) Biometrical analysis and selection of tetraploid progenies of *Panicum maximum* using mixed model methods. *Pesqui Agropecuária Bras* 39:335–341. doi: 10.1590/S0100-204X2004000400006
- Resende MAV de, Freitas JA de, Lanza MA, et al (2014) Divergência genética e índice de seleção via BLUP em acessos de algodoeiro para características tecnológicas da fibra. *Pesqui Agropecuária Trop* 44:334–640. doi: 10.1590/S1983-40632014000300006
- Resende MDV de (2002) *Genética biométrica e estatística no melhoramento de plantas perenes*. Brasília: Embrapa Informação Tecnológica
- Resende MDV de (2016) Software Selegen-REML/BLUP: a useful tool for plant breeding. *Crop Breed Appl Biotechnol* 16:330–339. doi: 10.1590/1984-70332016v16n4a49
- Resende MDV de (2007) *Matemática e Estatística na Análise de Experimentos e no Melhoramento Genético*, 1. ed. Colombo-PR
- Resende MDV de (2015) *Genética Quantitativa e de Populações*, 1st edn. viçosa
- Resende MDV de, Duarte JB (2007) Precisão e controle de qualidade em experimentos de avaliação de cultivares. *Pesqui Agropecuária Trop* 37:182–194.
- Sera T (2001) Coffee Genetic Breeding at IAPAR. *Crop Breed Appl Biotechnol* 1:179–199. doi: 10.13082/1984-7033.v01n02a08
- Silva FL da, Baffa DCF, Rezende JC de, et al (2015) Variabilidade genética entre genótipos de café robustas no estado de minas gerais. *Coffee Sci* 10:20–27.
- 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. *Ind Crops Prod* 69:355–361. doi: 10.1016/j.indcrop.2015.02.024
- Vencovsky R, Barriga P (1992) *Genética biométrica no fitomelhoramento*. Sociedade Brasileira de Genética, Ribeirão Preto
- Verardi CK, Scaloppi Junior EJ, Gonçalves P de S (2011) Prediction of direct and indirect genetic gains and genotypic correlations in rubber tree progenies. *Pesqui Agropecuária Bras* 46:1076–1081. doi: 10.1590/S0100-204X2011000900015

CAPÍTULO 3

Seleção genômica ampla em *Coffea canephora*

RESUMO

Os programas de melhoramento genético têm buscado, ao longo dos anos, estratégias eficientes para seleção de genótipos que apresentem desempenho superior. A seleção genômica ampla (GWS) surgiu em 2001 com o objetivo de aumentar a eficiência e acelerar o ganho de seleção. Essa técnica é considerada ferramenta essencial no melhoramento genético de espécies perenes, como *Coffea canephora*, devido principalmente, ao seu potencial em aumentar o ganho por unidade de tempo. Dessa forma, o objetivo do trabalho foi aplicar o princípio da GWS e avaliar sua eficiência em população de *C. canephora*, utilizando marcadores moleculares SNP e avaliações de oito características fenotípicas. As características fenotípicas avaliadas foram vigor vegetativo, incidência à ferrugem, incidência de cercosporiose, época de maturação dos frutos, tamanho de fruto, altura da planta, diâmetro da projeção da copa e produção por planta. Foram avaliados 165 genótipos, sendo 51 genótipos Conilon, 32 Robusta e 82 híbridos interpopulacionais. Utilizando a plataforma de sequenciamento da empresa RAPiD Genomics foram identificados 18.111 marcadores SNP, restando após análises de qualidade 14.429 marcadores. Foi observado boa capacidade preditiva para as características, exceto para época de maturação dos frutos, tamanho de fruto e produção por planta. Os menores valores de herdabilidade genômica encontrados para essas características podem justificar os baixos valores de capacidade preditiva obtidos. Os valores de acurácia estimados foram considerados de magnitude moderada a alta, variando entre 67% a 82%. Com a redução do tempo necessário para completar o ciclo seletivo de seis para três anos, a GWS proporcionou eficiência seletiva variando entre 22% a 146%. Esses resultados indicam que é possível alcançar maiores ganhos por unidade de tempo por meio da GWS. Portanto, a GWS se mostrou ser ferramenta útil para o melhoramento genético de *C. canephora*, por predizer com acurácia os fenótipos dos indivíduos e possibilitar a redução no tempo necessário para completar o ciclo de seleção, proporcionando ganhos em eficiência seletiva por unidade de tempo.

Palavras-chave: Café, GBLUP, capacidade preditiva, acurácia seletiva.

Introdução

Nos programas de melhoramento genético, por muito tempo, selecionou-se os melhores indivíduos por meio de marcadores morfológicos. Estes são influenciados pelo ambiente e, para muitas características, apresentam baixo ganho com seleção (Toppa and Jadoski 2013). Com o aperfeiçoamento das técnicas de biologia molecular, o uso de marcadores moleculares possibilitou avanços significativos na seleção de indivíduos superiores por permitir a detecção do polimorfismo ao nível do DNA (Sousa et al. 2017; Alkimim et al. 2017).

No início da década de 90, a Seleção Assistida por Marcadores Moleculares (SAM) foi proposta, a qual se fundamenta na existência de desequilíbrio de ligação entre marcadores de DNA e genes de interesse (Lande and Thompson 1990). Ao se comparar a SAM, para características oligogênicas, com a seleção fenotípica, verifica-se, entre outros benefícios, aumento na eficiência seletiva e redução do tempo necessário para praticar seleção (Noir et al. 2003; Gartner et al. 2013; Romero et al. 2014; Alkimim et al. 2017). Além disso, essa técnica possibilitou o melhoramento preventivo por permitir a seleção de genótipos portadores de genes de interesse em regiões onde a patógeno está ausente (Alkimim et al. 2017). No entanto, essa técnica tem se mostrado mais eficiente para características monogênicas ou oligogênicas e com alta herdabilidade (Asins et al. 2012; Kemper and Goddard 2012).

No geral, as características agronômicas são governadas por muitos genes comprometendo a eficiência da SAM. Para contornar essa limitação, um novo método de seleção, denominado Seleção Genômica Ampla (*Genome Wide Selection - GWS*) foi desenvolvido (Meuwissen et al. 2001). A GWS enfatiza a predição simultânea dos efeitos genéticos de centenas ou milhares de marcadores que cobrem densamente o genoma. Assim, é esperado que todos os genes ou QTL (*Quantitative Trait Loci*) de um caráter quantitativo estejam em desequilíbrio de ligação com pelo menos uma parte dos marcadores (Valente et al. 2016). A GWS se destaca por apresentar alta acurácia seletiva e por não exigir o prévio conhecimento da localização (mapas) dos QTL nos cromossomos (Meuwissen 2007; Jannink et al. 2010; Almeida et al. 2016).

Na abordagem da GWS, os valores genéticos genômicos dos indivíduos (*Genomic Estimated Breeding Value - GEBV*) podem ser preditos por diferentes metodologias estatísticas, entre as quais pode-se citar o GBLUP (*Genomic Best Linear Unbiased Prediction*) (VanRaden 2008). Nesse método, os valores de GEBV são preditos por meio da matriz de parentesco, estimada a partir de informações de marcadores moleculares, denominada de matriz de parentesco genômica. É assumido que cada marcador explica igual proporção da variância genética, ou seja, estima os GEBV considerando efeito poligênico em que todos marcadores apresentam variância constante (VanRaden 2008). O GBLUP apresenta como vantagens sua relativa simplicidade e

menor tempo computacional exigido (Heslot et al. 2015). Esse método é recomendado principalmente para características poligênicas, as quais são governadas por muitos genes de efeito menor (VanRaden 2008).

Dentre os marcadores moleculares que podem ser utilizados em estudos de GWS, os marcadores SNP (*Single Nucleotide Polymorphism*) têm se destacado por serem o tipo de polimorfismo mais abundante nos genomas, pela possibilidade de automatização, serem codominantes e bialélicos (Resende et al. 2008; Liao and Lee 2010). Recentemente, com o desenvolvimento das plataformas de NGS (*Next Generation Sequencing*), a descoberta dos SNP foi facilitada e os custos por *data point* reduzidos. Com a identificação de marcadores SNP amplamente distribuídos no genoma da espécie, a GWS tornou-se realidade e tem permitido ganhos significativo para diversos programas de melhoramento (Goddard and Hayes 2007; Meuwissen 2007; Carvalho and Silva 2010; Resende et al. 2012b; Fritsche-Neto et al. 2012; Van Eenennaam et al. 2014; Zhao et al. 2015).

Embora apresente importância econômica mundial, até o presente momento, nenhum estudo de GWS para *Coffea canephora* foi publicado. Essa espécie se caracteriza por ser alógama, diploide ($2n=2x=22$) e apresenta sistema de autoincompatibilidade gametofítica (Leroy et al. 2005). A espécie *C. canephora* se destaca por sua rusticidade, elevado potencial produtivo, maior conteúdo de sólidos solúveis totais e por apresentar resistência genética ao agente causal da principal doença do cafeeiro, a ferrugem alaranjada causada pelo fungo *Hemileia vastatrix* (Zambolim 2016).

Diante do exposto, objetivou-se aplicar o princípio da GWS e avaliar sua eficiência em *C. canephora*, por meio de marcadores moleculares SNP e avaliações de características fenotípicas.

Material e métodos

Material genético

A população de estudo foi composta por clones dos grupos varietais Conilon e Robusta e por híbridos interpopulacionais originados de cruzamentos entre estes grupos. O material genético de Conilon foi obtido do Instituto Capixaba de Pesquisa, Assistência Técnica e Extensão Rural (INCAPER) e o material de Robusta foi obtido do Centro Agronômico Tropical de Investigación y Enseñanza (CATIE). Essa população compõem o programa de melhoramento da Empresa de Pesquisa Agropecuária de Minas Gerais (Epamig) em parceria com a Universidade Federal de Viçosa (UFV) e a Empresa Brasileira de Pesquisa Agropecuária - Café (Embrapa Café), localizado no município de Oratórios/MG.

Os grupos varietais Conilon e Robusta foram constituídos por 51 e 32 genótipos (Tabela 1), respectivamente. Além desses genótipos, 82 híbridos interpopulacionais foram obtidos por meio de cruzamentos artificiais entre cinco genótipos do grupo Conilon (genitores masculinos) e cinco do grupo Robusta (genitores femininos), avaliados em dialelo parcial interpopulacional (Tabela 2).

Avaliações fenotípicas

As avaliações fenotípicas foram realizadas para oito características durante três anos consecutivos (2014 a 2016). Foram avaliadas cinco características categóricas e três características contínuas. Estas foram realizadas na época de maturidade fisiológica dos frutos dos cafeeiros.

As características categóricas avaliadas foram: vigor vegetativo (Vig), avaliação das incidências, em campo, de ferrugem (Fer) e de cercosporiose (Cer), época de maturação dos frutos (Mat) e tamanho de fruto (TF). O vigor vegetativo foi avaliado pelo aspecto geral da planta, observando-se o enfolhamento, a coloração das folhas, o estado nutricional e a sanidade dos cafeeiros. A escala de notas de 1 a 10 foi utilizada, sendo a nota 1 atribuída às plantas totalmente depauperada e a 10 às plantas altamente vigorosas; a incidência à ferrugem do cafeeiro, causada pelo fungo *Hemileia vastatrix* Berk. & Br., foi avaliada com notas de 1 a 5. A nota 1 foi atribuída para genótipos sem nenhum sintoma do patógeno e a nota 5 para genótipos altamente suscetíveis; a incidência à cercosporiose, causada pelo fungo *Cercospora coffeicola*, foi avaliada por notas de 1 a 5. A nota 1 foi atribuída para genótipos sem nenhum sintoma do patógeno e a nota 5 para genótipos altamente suscetíveis; a época de maturação dos frutos do cafeeiro foi classificada em precoce, média e tardia, recebendo notas de 1 a 3, respectivamente; e o tamanho dos frutos foi classificado como pequeno, médio e graúdo, notas 1, 2 e 3, respectivamente.

As características contínuas avaliadas foram: altura da planta (AP), diâmetro da projeção da copa (DC) e produção em litros por planta (Prod). A altura da planta, em centímetros (cm), foi determinada pela medida da ramificação ortotrópica mais desenvolvida, do nível do solo até o último ponto apical do cafeeiro com o auxílio de uma trena métrica afixada em uma haste de madeira; o diâmetro da projeção da copa da planta foi determinado em centímetros (cm) por meio de régua no sentido perpendicular à linha de plantio; e a produção por planta de cafeeiros foi avaliada colhendo todos os frutos presentes em um genótipo e mensurado o volume total em litros de café, recém colhido.

Análise dos dados fenotípicos

Os fenótipos foram corrigidos para efeitos ambientais de anos e blocos, utilizando o *software* Selegen REML/BLUP (Resende 2016).

A acurácia seletiva, baseada nos dados fenotípicos, foi obtida a partir da metodologia de modelos mistos do tipo REML/BLUP, também denominada metodologia de modelos mistos (Henderson 1975), utilizando o *software* Selegen REML/BLUP (Resende 2016). O modelo utilizado equivale a: $y = Xu + Tc + Wf + Zm + Qs + Sb + e$, em que, y é o vetor de dados; u é o vetor de efeitos de médias de anos (assumidos como fixos) somados à média geral; c é o vetor de efeitos da capacidade específica de combinação entre os genitores Conilon e Robusta (assumidos como aleatórios); f é o vetor de efeitos da capacidade geral de combinação do parental Robusta (assumidos como aleatórios); m é o vetor de efeitos da capacidade geral de combinação do parental Conilon (assumidos como aleatórios); s é o vetor de efeitos de ambiente permanente de indivíduos (assumidos como aleatórios); b é o vetor de efeitos de ambiente permanente de blocos (assumidos como aleatórios); e é o vetor de resíduos (assumidos como aleatórios). As letras maiúsculas representam as matrizes de incidência para os referidos efeitos.

A acurácia foi obtida por meio da equação: $\hat{r}_{gg} = (1 - PEV/\sigma_g^2)^{1/2}$, em que σ_g^2 é a variação genotípica entre os materiais em avaliação e PEV é a variância do erro de predição, dada por $PEV = C_i^{22} \sigma_e^2$, em que C_i^{22} = i-ésimo elemento da diagonal da inversa da matriz dos coeficientes das equações de modelo misto e σ_e^2 é a variância residual.

Extração do DNA genômico, identificação e análise de qualidade de SNP

Folhas jovens e completamente expandidas dos 165 cafeeiros em estudo 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 do DNA das amostras foi padronizada e as mesmas enviadas para a empresa RAPiD Genomics localizada na Florida/EUA, para construção das sondas, sequenciamento e identificação dos marcadores moleculares SNP.

Para identificação dos marcadores SNP e genotipagem dos cafeeiros foram utilizadas 10.000 sondas (Resende et al. 2016). As sondas foram construídas a partir de sequências de referências. Um dos bancos de dados utilizado foi o Projeto Brasileiro do Genoma Café, que contém mais de 200 mil ESTs (*Expressed Sequence Tag*), correspondendo a aproximadamente 33 mil genes transcritos, chamados de Unigenes (Vieira et al. 2006). O outro banco usado foi o genoma da espécie *C. canephora*, contendo um total de 25.574 genes (Denoëud et al. 2014). Por

meio dessas sequências de referências, sondas específicas foram obtidas, de modo que todo o genoma fosse coberto, considerando regiões transcritas e não transcritas. Com essas sondas, os genótipos foram sequenciados via plataforma Illumina e os marcadores SNP identificados por meio de ferramentas de bioinformática (Resende et al. 2016).

O controle de qualidade dos SNP foi realizado por meio da eliminação de locos pouco polimórficos ($MAF - \text{minor allele frequency} = 5\%$); e/ou dos locos com baixa taxa de atendimento na genotipagem dos indivíduos ($\text{call rate} = 95\%$). O nível crítico para o parâmetro MAF foi obtido por meio da equação $MAF = \frac{1}{\sqrt{2N}}$, onde N refere-se ao número de genótipos avaliados (Resende et al. 2017).

Predição utilizando o modelo GBLUP

As análises foram realizadas utilizando o método GBLUP via procedimento RKHS (*Reproducing Kernel Hilbert Spaces*) (Gianola 2006) com algoritmo Bayesiano, via ambiente R, no pacote BGLR (Perez and De Los Campos 2014).

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 corrigidas, 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 W 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:

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

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 .

Validação cruzada

O método de validação cruzada empregado foi o *K-Fold*, sendo considerado nesse trabalho k igual a 11 Folds. O conjunto de observações foi dividido aleatoriamente em grupos. No processo

de análise, 150 genótipos eram utilizados como população de treinamento, e o grupo de 15 genótipos restantes eram utilizados como população de validação. Esse procedimento foi repetido por 11 vezes ($k=11$) e todos os grupos de genótipos excluídos foram utilizados na validação.

Capacidade preditiva, viés da predição, acurácia da GWS e n_{QTL}

A capacidade preditiva e o viés da predição são medidas práticas da capacidade do método fazer predições de forma acurada e não viesada. A capacidade preditiva (r_{gy}) é obtida por meio da correlação entre os valores genômicos preditos e os valores fenotípicos observados, sendo equivalente à capacidade preditiva da GWS em estimar os fenótipos. O viés da predição (b) é obtido por meio do coeficiente de regressão dos valores genômicos preditos sobre os valores fenotípicos (Resende et al. 2012b; Pértile et al. 2016).

A acurácia foi obtida pelo estimador: $r_{gg} = r_{gy}/\sqrt{h^2}$, em que r_{gy} equivale a capacidade preditiva da GWS e h^2 a herdabilidade individual (Borém and Fritsche-Neto 2013).

A estimativa do número de QTL que controlam cada característica foi calculada 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. 2014).

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

Eficiência da GWS

A eficiência seletiva da GWS em comparação com a seleção baseada apenas em fenótipos de seis anos, foi calculada usando a expressão: $Ef = \frac{r_{gy}L_f}{r_{yy}L_{GWS}}$, em que r_{gy} é a capacidade preditiva da GWS; r_{yy} é a acurácia da seleção com base em fenótipos; L_f é o tempo médio requerido para o ciclo de seleção com base em fenótipos; e L_{GWS} é o tempo médio requerido para o ciclo de seleção com base na GWS (Resende et al. 2012a).

Resultados e discussão

Com o sequenciamento dos 165 genótipos utilizando 10.000 sondas, distribuídas ao longo de todo o genoma, foram identificados 18.111 SNP. Após as análises de qualidade, foram obtidos

14.429 marcadores SNP. As avaliações de qualidade dos SNP permitem a identificação de marcadores com os critérios de qualidade estabelecidos como ideais. As análises de qualidade são vantajosas por remover os marcadores de baixa qualidade antes das análises estatísticas e, dessa forma, reduz a ocorrência de falso-positivos (erro tipo I) e falso-negativos (erro tipo II) (Anderson et al. 2010). Alta densidade de marcadores é importante para capturar genes de efeito menor e maior e, conseqüentemente, aumentar a probabilidade de explicar a maior parte da variação genética do caráter em estudo (Resende et al. 2008; Resende et al. 2016). Em trabalho realizado com dados simulados, foi verificado que o uso de maiores densidades de marcadores é necessário, para que se obtenha acurácia de predição de alta magnitude ($> 70\%$) (Valente et al. 2016).

As estimativas dos valores de herdabilidade genômica, capacidade preditiva da GWS, viés da predição, acurácia com base nos dados fenótipos, acurácia da GWS e número de QTLs que controlam a característica estão apresentadas na Tabela 3.

Os valores estimados de herdabilidade genômica (h_a^2) variaram de 0,15 para a característica produção por planta (Prod) a 0,53 para a característica diâmetro da projeção da copa (DC). Esses resultados indicam que a seleção genômica é herdável. Embora, valores consideráveis de h_a^2 tenham sido obtidos para as características época de maturação dos frutos (0,21), tamanho de fruto (0,21) e produção por planta (0,15), não foi possível obter boa capacidade preditiva para essas características. Espera-se que as estimativas de capacidade preditiva sejam menores para as características que apresentam baixa herdabilidade (Legarra et al. 2008). Nesse sentido, como os menores valores de h_a^2 foram obtidos para essas três características, pode-se justificar o fato de não terem sido obtidos bons valores de capacidade preditiva para tais características em questão.

A capacidade preditiva (r_{gy}) equivale a correlação entre os valores genéticos genômicos e os valores fenotípicos observados, e reflete a consistência das informações moleculares em informar sobre o fenótipo (Cavalcanti et al. 2012). Foi observado boa capacidade preditiva para as características Vig (0,44), Fer (0,48), Cer (0,54), AP (0,41) e DC (0,58), indicando a capacidade de se antever fenótipos para essas características. Esses dados mostram que, de modo geral, os valores de r_{gy} foram maiores para as características que apresentaram os maiores valores de h_a^2 . Assim como observado nesse trabalho, foi verificado resposta da capacidade preditiva em função da herdabilidade em estudo de GWS com cajueiro (*Anacardium occidentale*) (Cavalcanti et al. 2012).

O viés da predição (b), obtido por meio do coeficiente de regressão dos valores fenotípicos sobre os valores genômicos preditos, apresentou valores próximos de 1,0 para as características vigor vegetativo, incidência de ferrugem, incidência de cercosporiose, altura da planta, diâmetro da projeção da copa, época de maturação dos frutos e tamanho de fruto. Valores de viés próximos

de 1,0 indicam que a predição foi não-viesada e, portanto, são efetivas em predizer as reais magnitudes das diferenças entre os indivíduos avaliados (Resende et al. 2012b). A característica produção por planta não apresentou viés da predição próximo da unidade, indicando para essa característica que a predição foi viesada. Essa característica apresentou o menor valor de estimativa de herdabilidade genômica, sendo uma possível justificativa para o viés observado. Além disso, características governadas por maior quantidade de genes exigem populações com maior tamanho amostral.

A acurácia da GWS (r_{gg}) reflete a quantidade e qualidade das informações e procedimentos utilizados na predição dos valores genéticos genômicos dos indivíduos. Ela se baseia na correlação entre o valor genético verdadeiro e aquele estimado ou predito (Rabier et al. 2016). A acurácia pode ser classificada em muito alta, quando maior que 90%; alta, quando entre 70% e 90%; moderada, entre 50% e 70% e baixa, quando menor que 50% (Resende and Duarte 2007). Para as características época de maturação dos frutos, tamanho de fruto e produção por planta não foi verificada capacidade preditiva satisfatória e, portanto, para tais características não foram estimadas as acurácias. Para as demais características as estimativas dos valores de acurácia foram obtidas. Os valores de acurácia observados nesse trabalho foram considerados de magnitude moderada a alta, variando entre 67% (Vig) a 82% (Fer). Valores moderados a altos de r_{gg} , 68 e 79%, foram obtidos até mesmo para as características altura da planta e incidência de ferrugem, que apresentaram menores valores de h_a^2 , 0,36 e 0,37, respectivamente. Esses resultados confirmam a eficiência da GWS na seleção de características que apresentam baixa herdabilidade, estando de acordo com outros trabalhos (Legarra et al. 2008; Zhang et al. 2010).

O número estimado de QTLs que controlam cada característica variou entre 35 (Cer) a 87 (Vig). Por meio desses resultados é possível verificar a natureza quantitativa das características avaliadas nesse estudo. Além disso, observa-se que os menores valores de acurácia da GWS, 67% (Vig) e 68% (AP), foram obtidos para as características que apresentaram maior número de QTLs. Em trabalho realizado com dendê (*Elaeis guineensis* Jacq.) foi também verificado que a acurácia da GWS é inversamente proporcional ao número de QTLs que controlam a característica (Wong and Bernardo 2008). Isso é esperado porque quando a característica é governada por maior número de QTLs, mais complexa ela se torna. Além disso, em características poligênicas, em geral, verifica-se que cada gene contribui com pequeno efeito para a manifestação do caráter. No entanto, mesmo em cenário com baixa herdabilidade e maior número de QTL, os resultados foram promissores. Isso se deve, provavelmente, ao conjunto de SNP selecionados pelos critérios de qualidade e, sobretudo, por serem amplamente distribuídos em todo o genoma dessa espécie.

Na Tabela 4 está apresentado a estimativa do número necessário de indivíduos que devem ser avaliados para alcançar acurácia desejada. O número de indivíduos foi obtido considerando valores de acurácia desejada de 0,50, 0,60, 0,70, 0,80 e 0,90. Nesse cálculo foram considerados as estimativas dos valores de herdabilidade genômica (h_a^2) e o número estimado de QTLs (n_{QTL}) que controlam cada característica, apresentados na Tabela 3.

Para se obter acurácia desejada (r_{ggd}) de 70%, valor considerado de alta magnitude (Resende and Duarte 2007), seria necessário avaliar 194 indivíduos para a característica vigor vegetativo, 96 para incidência à ferrugem do cafeeiro, 78 para incidência de cercosporiose, 184 para altura da planta e 89 para diâmetro da projeção da copa. Dessa forma, observa-se que para a maioria das características, Fer, Cer e DC, foram avaliados neste trabalho mais indivíduos do que o necessário para alcançar acurácia de 70%. Também foi possível verificar para todas as características que quanto maior a acurácia desejada, maior número de indivíduos deve ser analisado.

Na Figura 1 está apresentada a eficiência da GWS com a redução do ciclo seletivo em relação a seleção baseada somente em dados fenotípicos, de seis anos, para todas características que apresentaram boa capacidade preditiva, exceto incidência de cercosporiose (Cer). Para essa variável a herdabilidade no sentido amplo obtida com base nos dados fenotípicos no *software* Selegen REML/BLUP foi igual a zero e, portanto, não foi incluída na análise de eficiência da GWS. Dessa forma, foi possível estimar a eficiência da GWS somente para as características vigor vegetativo, incidência de ferrugem, altura da planta e diâmetro da projeção da copa. Em espécies perenes, como *C. canephora*, uma das vantagens da seleção genômica é a redução do ciclo de seleção ao praticar a seleção precoce (Castro et al. 2016).

Com base na Figura 1 é possível verificar o aumento da eficiência seletiva com o uso da GWS, mesmo para características como vigor vegetativo e altura da planta, as quais apresentaram altos valores de estimativa de acurácia a partir da seleção baseada em dados fenotípicos, 60 e 67%, respectivamente (Tabela 2). Esse aumento da eficiência se deve a redução do tempo necessário para completar um ciclo seletivo com o uso da GWS. Dessa forma, com a redução do ciclo, variando de seis a três anos, tem-se um aumento na eficiência seletiva da GWS para todas as características ao longo da redução do ciclo. Portanto, mesmo quando a acurácia da seleção genômica seja de mesma magnitude que a obtida com a seleção baseada em dados fenotípicos, a GWS proporcionará ganhos genéticos superiores, devido a redução do ciclo de seleção (Gois et al. 2016).

Com a redução do ciclo seletivo de seis para três anos, a GWS foi superior (variando entre 22% a 146%) para todas as características. Dessa forma, observa-se que a redução no tempo

necessário para completar um ciclo seletivo é expressiva com a GWS. Isso se deve ao fato que a predição genômica e a seleção podem ser feitas no estágio de plântulas, e portanto, a GWS é mais eficiente por unidade de tempo (Resende et al. 2012b; Gois et al. 2016). Resultados semelhantes foram observados em outros trabalhos. Em trabalho avaliando a eficiência seletiva da GWS com redução de 50% no ciclo seletivo do *Citrus*, a GWS foi superior para todas as características avaliadas (variando entre 31% a 160%) (Gois et al. 2016). Também foi verificado em trabalho conduzido com dendê (*Elaeis guineensis* Jacq.), a redução do ciclo de seleção de 19 para seis anos com o uso da GWS (Wong and Bernardo 2008). Em trabalho conduzido com milho (*Zea mays* L.) foi observado que com o uso da GWS houve aumento significativo na acurácia seletiva e nos ganhos genéticos por unidade de tempo (Fritsche-Neto et al. 2012). Em outros trabalhos, a seleção genômica também revelou o grande potencial do seu uso em aumentar a eficiência do melhoramento, via resultados de dados simulados (Resende et al. 2008; Valente et al. 2016).

Mesmo para as características que apresentaram menores valores de h_a^2 , altura da planta (0,36) e incidência de ferrugem (0,37), foram obtidos ganhos em eficiência com a GWS. Esses dados confirmam a importância da GWS na seleção de características que apresentam baixa herdabilidade, fato este também observado em outros trabalhos (Resende et al. 2012b; Gois et al. 2016).

Conclusão

Os dados obtidos nesse trabalho demonstram que a seleção genômica ampla é ferramenta útil para o melhoramento de *C. canephora*, por prever com acurácia os fenótipos dos indivíduos e possibilitar a redução no tempo necessário para completar o ciclo de seleção, proporcionando ganhos em eficiência seletiva por unidade de tempo.

Tabela 1. Genótipos do grupo varietal Conilon

Grupo varietal	Genótipos			
Conilon	UFV 513	UFV 3628-23	UFV 3628-45	UFV 3629-31
	UFV 3627-20	UFV 3628-24	UFV 3628-46	UFV 3629-34
	UFV 3627-24	UFV 3628-26	UFV 3628-5	UFV 3629-4
	UFV 3627-27	UFV 3628-27	UFV 3629-10	UFV 3630-10
	UFV 3627-29	UFV 3628-28	UFV 3629-11	UFV 3630-2
	UFV 3627-30	UFV 3628-29	UFV 3629-17	UFV 3630-5
	UFV 3627-31	UFV 3628-3	UFV 3629-24	UFV 3630-6
	UFV 3628-1	UFV 3628-32	UFV 3629-25	UFV 3630-7
	UFV 3628-10	UFV 3628-35	UFV 3629-26	UFV 3631-10
	UFV 3628-16	UFV 3628-4	UFV 3629-27	UFV 3631-11
	UFV 3628-17	UFV 3628-40	UFV 3629-28	UFV 3631-13
	UFV 3628-2	UFV 3628-42	UFV 3629-29	UFV 3631-9
	UFV 3628-22	UFV 3628-44	UFV 3629-30	-
Robusta	UFV 3356-71	UFV 3362-118	UFV 3367-101	UFV 3373-43
	UFV 3356-76	UFV 3363-125	UFV 3367-105	UFV 3374-29
	UFV 3357-93	UFV 3365-144	UFV 3368-52	UFV 3375-65
	UFV 3358-88	UFV 3366-134	UFV 3368-58	UFV 3375-66
	UFV 3360-169	UFV 3366-138	UFV 3370-50	UFV 3376-8
	UFV 3360-171	UFV 3366-139	UFV 3371-19	UFV 3376-9
	UFV 3361-148	UFV 3367-96	UFV 3371-20	UFV 3377-12
	UFV 3361-151	UFV 3367-97	UFV 3373-36	UFV 514

Tabela 2. Híbridos e a descrição dos cruzamentos para obtenção dos mesmos

Híbridos	Cruzamentos
H092-2, H092-8, H092-10, H092-20, H092-22, H092-35	UFV 3366-139 x UFV 513
H093-1, H093-10, H093-12, H093-20	UFV 3366-139 x UFV 3629-11
H094-2, H094-5, H094-12, H094-13, H094-18, H094-30	UFV 3367-98 x UFV 513
H095-9, H095-11, H095-14, H095-25, H095-28	UFV 3366-139 x UFV 3627-31
H097-8, H097-13, H097-16, H097-17, H097-18, H097-19	UFV 3374-28 x UFV 3627-31
H098-1, H098-2, H098-3, H098-5, H098-13	UFV 3365-144 x UFV 3629-11
H099-1, H099-2, H099-7, H099-8, H099-10	UFV 3365-144 x UFV 513
H0910-4, H0910-5, H0910-6, H0910-7, H0910-8	UFV 3367-98 x UFV 3629-11
H0911-2, H0911-3, H0911-4, H0911-5, H0911-6, H0911-7	UFV 3367-98 x UFV 3628-2
H0912-1, H0912-3, H0912-4, H0912-6, H0912-7, H0912-8	UFV 3366-139 x UFV 3628-2
H0913-1, H0913-3, H0913-4, H0913-5, H0913-7	UFV 3365-144 x UFV 3628-2
H0914-2, H0914-3, H0914-4, H0914-5, H0914-6	UFV 3374-28 x UFV 513
H0915-1, H0915-2, H0915-3, H0915-4, H0915-5	UFV 3373-36 x UFV 3628-2
H0916-1, H0916-2, H0916-4	UFV 3367-98 x UFV 3629-25
H0917-1, H0917-2, H0917-3	UFV 3365-144 x UFV 3629-25
H0918-1, H0918-2, H0918-3	UFV 3373-36 x UFV 513
H0919-2	UFV 3373-36 x UFV 3629-11
H0920-1	UFV 3373-36 x UFV 3629-25
H0921-1	UFV 3374-28 x UFV 3629-11
H0922-1	UFV 3373-36 x UFV 3627-31

Tabela 3. Resultados da Seleção Genômica Ampla (GWS) para oito características morfoagronômicas em população de melhoramento de *Coffea canephora*

Características	$h_a^2 \pm sd$	$r_{gy} \pm sd$	$b \pm sd$	r_{yy}	r_{gg}	nQTL
Vig	0,43 \pm 0,04	0,44 \pm 0,15	1,18 \pm 0,50	0,60	0,67	87
Fer	0,37 \pm 0,04	0,48 \pm 0,15	1,40 \pm 0,67	0,39	0,79	37
Cer	0,43 \pm 0,04	0,54 \pm 0,20	1,31 \pm 0,57	-	0,82	35
AP	0,36 \pm 0,04	0,41 \pm 0,22	1,35 \pm 0,86	0,67	0,68	69
DC	0,53 \pm 0,03	0,58 \pm 0,13	1,23 \pm 0,41	0,57	0,80	49
Mat	0,21 \pm 0,19	-0,03 \pm 0,19	1,08 \pm 2,99	-	-	-
TF	0,21 \pm 0,19	0,00 \pm 0,29	0,60 \pm 1,42	-	-	-
Prod	0,15 \pm 0,02	-0,02 \pm 0,30	-0,11 \pm 2,12	-	-	-

h_a^2 : herdabilidade genômica; sd: desvio padrão; r_{gy} : Capacidade preditiva; b: viés da predição; r_{yy} : acurácia da seleção com base nos dados fenótipos obtidos pelo método REML/BLUP; r_{gg} : acurácia da GWS; nQTL: número de QTLs que controlam a característica; Vig: vigor vegetativo; Fer: incidência de ferrugem; Cer: incidência de cercosporiose; AP: altura da planta; DC: diâmetro da projeção da copa; Mat: época de maturação dos frutos; TF: tamanho de fruto; Prod: produção por planta.

Tabela 4. Número de indivíduos necessários para se obter acurácia desejada da GWS (r_{ggd}) em população de melhoramento de *Coffea canephora* para as características Vig, Fer, Cer, AP e DC

r_{ggd}	Características				
	Vig	Fer	Cer	AP	DC
0,50	67	33	27	64	31
0,60	114	56	46	108	52
0,70	194	96	78	184	89
0,80	360	178	145	341	164
0,90	863	426	347	817	394

r_{ggd} : acurácia desejada da GWS; Vig: vigor vegetativo; Fer: incidência de ferrugem; Cer: incidência de cercosporiose; AP: altura da planta; DC: diâmetro da projeção da copa.

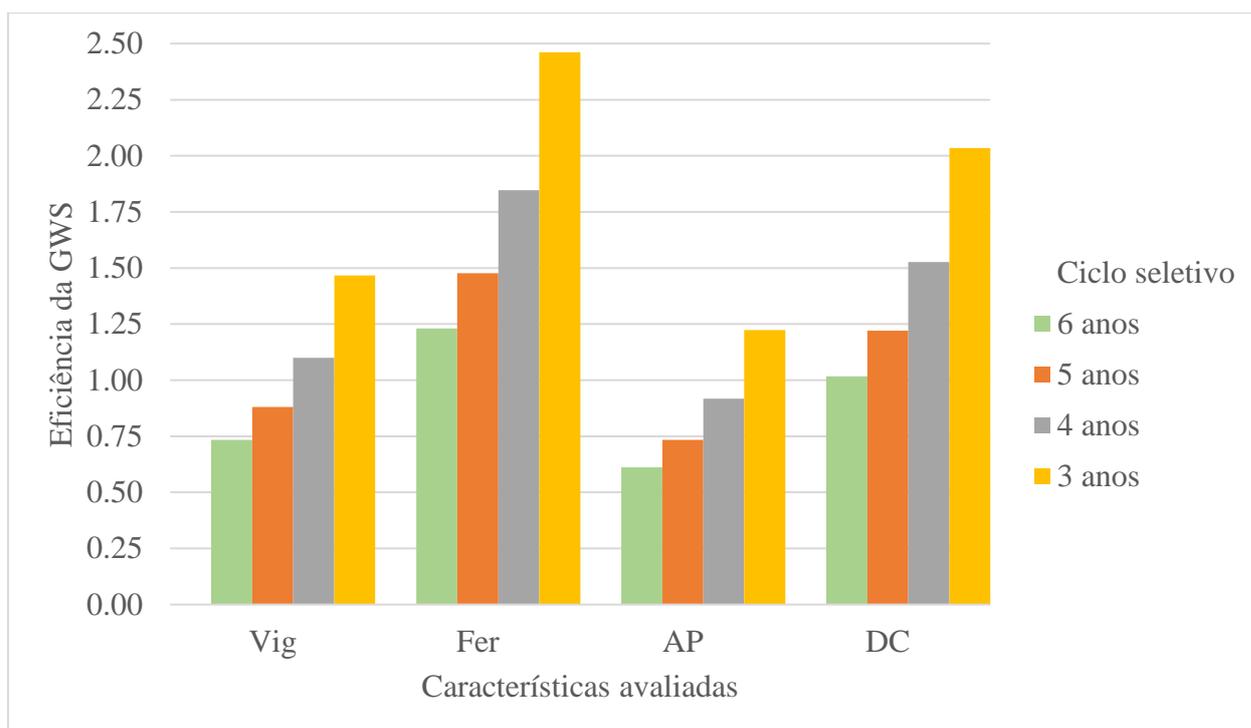


Figura 1. Eficiência da GWS em relação a seleção baseada somente em dados fenotípicos, de seis anos, em população de melhoramento de *Coffea canephora*, para as variáveis vigor vegetativo (Vig), incidência de ferrugem (Fer), incidência de cercosporiose (Cer), altura da planta (AP) e diâmetro da projeção da copa (DC)

Referências

- 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. *Mol Breed* 37:6. doi: 10.1007/s11032-016-0609-1
- Almeida ÍF de, Cruz CD, Resende MDV de (2016) Validação e correção de fenótipos na seleção genômica ampla. *Pesqui Agropecuária Bras* 51:1973–1982. doi: 10.1590/s0100-204x2016001200008
- Anderson CA, Pettersson FH, Clarke GM, et al (2010) Data quality control in genetic case-control association studies. *Nat Protoc* 5:1564–1573. doi: 10.1038/nprot.2010.116
- Asins MJ, Fernandez-Ribacoba J, Bernet GP, et al (2012) The position of the major QTL for Citrus tristeza virus resistance is conserved among *Citrus grandis*, *C. aurantium* and *Poncirus trifoliata*. *Mol Breed* 29:575–587. doi: 10.1007/s11032-011-9574-x
- Borém A, Fritsche-Neto R (2013) *Biotecnologia Aplicada ao Melhoramento de Plantas*.
- Carvalho MC da CG de, Silva DCG da (2010) Sequenciamento de DNA de nova geração e suas aplicações na genômica de plantas. *Ciência Rural* 40:735–744. doi: 10.1590/S0103-84782010000300040
- Castro CA de O, Resende RT, Bhering LL, Cruz CD (2016) Brief history of Eucalyptus breeding in Brazil under perspective of biometric advances. *Ciência Rural* 46:1585–1593. doi: 10.1590/0103-8478cr20150645
- Cavalcanti JJV, Resende MDV de, Santos FHC dos, Pinheiro CR (2012) Predição simultânea dos efeitos de marcadores moleculares e seleção genômica ampla em cajueiro. *Rev Bras Frutic* 34:840–846. doi: 10.1590/S0100-29452012000300025
- Denoeud F, Carretero-Paulet L, Dereeper A, et al (2014) The coffee genome provides insight into the convergent evolution of caffeine biosynthesis. *Science* (80-) 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 Breed Appl Biotechnol* 5:387–393. doi: 10.12702/1984-7033.v05n04a03
- Fritsche-Neto R, Resende MDV, Miranda GV, DoVale JC (2012) Seleção genômica ampla e novos métodos de melhoramento do milho. *Rev Ceres* 59:794–802. doi: 10.1590/S0034-737X2012000600009
- Gartner GAL, McCouch SR, Moncada MDP (2013) A genetic map of an interspecific diploid pseudo testcross population of coffee. *Euphytica* 192:305–323. doi: 10.1007/s10681-013-0926-y

- Gianola D (2006) Genomic-Assisted Prediction of Genetic Value With Semiparametric Procedures. *Genetics* 173:1761–1776. doi: 10.1534/genetics.105.049510
- Goddard ME, Hayes BJ (2007) Genomic selection. *J Anim Breed Genet* 124:323–330. doi: 10.1111/j.1439-0388.2007.00702.x
- Gois IB, Borém A, Cristofani-Yaly M, et al (2016) Genome wide selection in Citrus breeding. *Genet Mol Res* 15:1–14. doi: 10.4238/gmr15048863
- 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 Sci* 55:1–12. doi: 10.2135/cropsci2014.03.0249
- Jannink J-L, Lorenz AJ, Iwata H (2010) Genomic selection in plant breeding: from theory to practice. *Brief Funct Genomics* 9:166–177. doi: 10.1093/bfgp/elq001
- Kemper KE, Goddard ME (2012) Understanding and predicting complex traits: knowledge from cattle. *Hum Mol Genet* 21:R45–R51. doi: 10.1093/hmg/dds332
- Lande R, Thompson R (1990) Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics* 124:743–756.
- 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
- Leroy T, Marraccini P, Dufour M, et al (2005) Construction and characterization of a *Coffea canephora* BAC library to study the organization of sucrose biosynthesis genes. *Theor Appl Genet* 111:1032–1041. doi: 10.1007/s00122-005-0018-z
- Liao P-Y, Lee KH (2010) From SNPs to functional polymorphism: The insight into biotechnology applications. *Biochem Eng J* 49:149–158. doi: 10.1016/j.bej.2009.12.021
- Meuwissen T (2007) Genomic selection: marker assisted selection on a genome wide scale. *J Anim Breed Genet* 124:321–322. doi: 10.1111/j.1439-0388.2007.00708.x
- Meuwissen T, Hayes B, Goddard M (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157:1819–1829. doi: 11290733
- Noir S, Anthony F, Bertrand B, et al (2003) Identification of a major gene (Mex-1) from *Coffea canephora* conferring resistance to *Meloidogyne exigua* in *Coffea arabica*. *Plant Pathol* 52:97–103. doi: 10.1046/j.1365-3059.2003.00795.x
- 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
- Pértile SFN, Silva FF e, Salvian M, Mourão GB (2016) Seleção e associação genômica ampla para o melhoramento genético animal com uso do método ssGBLUP. *Pesqui Agropecuária Bras*

- 51:1729–1736. doi: 10.1590/s0100-204x2016001000004
- Rabier C-E, Barre P, Asp T, et al (2016) On the Accuracy of Genomic Selection. *PLoS One* 11:e0156086. doi: 10.1371/journal.pone.0156086
- 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, California, p 1
- Resende M De, Lopes P, Silva RL Da, Pires IE (2008) Seleção genômica ampla (GWS) e maximização da eficiência do melhoramento genético. *Pesqui Florest Bras* 56:63–77. doi: 51970036
- Resende MDV de (2016) Software Selegen-REML/BLUP: a useful tool for plant breeding. *Crop Breed Appl Biotechnol* 16:330–339. doi: 10.1590/1984-70332016v16n4a49
- Resende MDV de (2007) *Matemática e Estatística na Análise de Experimentos e no Melhoramento Genético*, 1. ed. Colombo-PR
- Resende MDV de, Duarte JB (2007) Precisão e controle de qualidade em experimentos de avaliação de cultivares. *Pesqui Agropecuária Trop* 37:182–194.
- Resende MDV de, Silva FF e, Azevedo CF (2014) *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 MDV de, Silva FF e, Ferreira C, et al (2017) Atualidades da biometria no melhoramento de plantas perenes. In: Ludke WH, Andrade ACB, Volpato L, et al. (eds) *Desafios Biométricos No Melhoramento Genético*, 1ª edição. Viçosa, p 166
- Resende MDV de, Silva FF e, Lopes PS, Azevedo CF (2012a) Seleção Genômica Ampla (GWS) via Modelos Mistos (REML/BLUP), Inferência Bayesiana (MCMC), Regressão Aleatória Multivariada e Estatística Espacial. 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 (2012b) Genomic selection for growth and wood quality in *Eucalyptus*: capturing the missing heritability and accelerating breeding for complex traits in forest trees. *New Phytol* 194:116–128. doi: 10.1111/j.1469-8137.2011.04038.x
- 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 Breed* 133:121–129. doi: 10.1111/pbr.12127
- Sousa TV, Caixeta ET, Alkimim ER, et al (2017) Molecular markers useful to discriminate *Coffea arabica* cultivars with high genetic similarity. *Euphytica*. doi: 10.1007/s10681-017-1865-9
- Toppa EVB, Jadoski CJ (2013) O uso dos marcadores moleculares no melhoramento genético de plantas. *Sci Agrar Parana* 12:1–5. doi: 10.18188/1983-1471
- Valente MSF, Viana JMS, Resende MDV de, et al (2016) Seleção genômica para melhoramento vegetal com diferentes estruturas populacionais. *Pesqui Agropecuária Bras* 51:1857–1867. doi: 10.1590/s0100-204x2016001100008
- Van Eenennaam AL, Weigel K a, Young AE, et al (2014) Applied Animal Genomics: Results from the Field. *Annu Rev Anim Biosci* 2:105–139. doi: 10.1146/annurev-animal-022513-114119
- VanRaden PM (2008) Efficient Methods to Compute Genomic Predictions. *J Dairy Sci* 91:4414–4423. doi: 10.3168/jds.2007-0980
- Vieira LGE, Andrade AC, Colombo CA, et al (2006) Brazilian coffee genome project: an EST-based genomic resource. *Brazilian J Plant Physiol* 18:95–108. doi: 10.1590/S1677-04202006000100008
- Wong CK, Bernardo R (2008) Genomewide selection in oil palm: increasing selection gain per unit time and cost with small populations. *Theor Appl Genet* 116:815–824. doi: 10.1007/s00122-008-0715-5
- Zambolim L (2016) Current status and management of coffee leaf rust in Brazil. *Trop Plant Pathol* 41:1–8. doi: 10.1007/s40858-016-0065-9
- Zhang Z, Liu J, Ding X, et al (2010) Best Linear Unbiased Prediction of Genomic Breeding Values Using a Trait-Specific Marker-Derived Relationship Matrix. *PLoS One* 5:e12648. doi: 10.1371/journal.pone.0012648
- Zhao Y, Mette MF, Reif JC (2015) Genomic selection in hybrid breeding. *Plant Breed* 134:1–10. doi: 10.1111/pbr.12231

Conclusões gerais

A metodologia de identificação dos marcadores SNP se mostrou promissora e os SNP obtidos foram eficientes nos estudos de diversidade e estrutura genética de genótipos da espécie *C. canephora*. Os maiores valores de distância genética foram obtidos entre genótipos dos grupos Conilon e Robusta como já se esperava, visto que estes consistem em dois grupos heteróticos dentro da espécie *C. canephora*.

Nas análises dos dados fenotípicos via modelos mistos (REML/BLUP), o estudo apresentou resultados de grande valia para programas de melhoramento genético do cafeeiro. Foram identificados genótipos promissores dentro de cada grupo heterótico, Conilon e Robusta, bem como famílias híbridas e genótipos Híbridos de interesse.

A análise de seleção genômica ampla, permitiu prever com acurácia os fenótipos dos indivíduos e a redução no tempo necessário para completar o ciclo de seleção, proporcionando ganhos em eficiência seletiva por unidade de tempo.