

UNIVERSIDADE FEDERAL DE VIÇOSA

LETÍCIA DE FARIA SILVA

**ESTRUTURA POPULACIONAL E DIVERSIDADE GENÉTICA DE *Coffea canephora*
DETECTADAS POR MARCADORES MOLECULARES**

**VIÇOSA-MINAS GERAIS
2022**

LETÍCIA DE FARIA SILVA

**ESTRUTURA POPULACIONAL E DIVERSIDADE GENÉTICA DE *Coffea canephora*
DETECTADAS POR MARCADORES MOLECULARES**

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*.

Orientadora: Eveline Teixeira Caixeta

**VIÇOSA – MINAS GERAS
2022**

**Ficha catalográfica elaborada pela Biblioteca Central da Universidade
Federal de Viçosa - Campus Viçosa**

T

S586e
2022
Silva, Letícia de Faria, 1992-
Estrutura populacional e diversidade genética de *Coffea
canephora* detectadas por marcadores moleculares / Letícia de
Faria Silva. – Viçosa, MG, 2022.
1 tese eletrônica (80 f.): il. (algumas color.).

Orientador: Eveline Teixeira Caixeta.

Tese (doutorado) - Universidade Federal de Viçosa,
Departamento de Fitotecnia, 2022.

Inclui bibliografia.

DOI: <https://doi.org/10.47328/ufvbbt.2022.602>

Modo de acesso: World Wide Web.

1. Café - Melhoramento genético. 2. Café - Mapeamento
genômico. 3. Café - Recursos do germoplasma. 4. Marcadores
genéticos. I. Caixeta, Eveline Teixeira, 1972-. II. Universidade
Federal de Viçosa. Departamento de Fitotecnia. Programa de
Pós-Graduação em Genética e Melhoramento. III. Título.

CDD 22. ed. 631.523

LETÍCIA DE FARIA SILVA

**ESTRUTURA POPULACIONAL E DIVERSIDADE GENÉTICA DE *Coffea canephora*
DETECTADAS POR MARCADORES MOLECULARES**

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: 12 de agosto de 2022

Assentimento:

Letícia de Faria Silva
Autora

Eveline Teixeira Caixeta
Orientadora

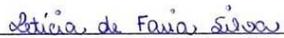
LETÍCIA DE FARIA SILVA

**ESTRUTURA POPULACIONAL E DIVERSIDADE GENÉTICA DE *Coffea canephora*
DETECTADAS POR MARCADORES MOLECULARES**

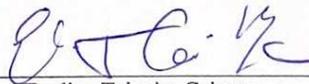
Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento, para obtenção do título de *Doctor Scientiae*.

APROVADA: 12 de agosto de 2022

Assentimento:



Letícia de Faria Silva
Autora



Eveline Teixeira Caixeta
Orientadora

À minha mãe Esméria, ao Bruno
E a toda a minha família,

Dedico.

AGRADECIMENTOS

A Deus pela sabedoria, calma e, sobretudo, sua benção concedida a mim nessa trajetória.

À minha mãe, Esméria, por ser minha maior incentivadora e expectadora, acreditando mais em mim do que eu mesma.

Ao meu pai, Naelson, pelo apoio e conselhos.

As minhas avós, Rita e Luzia, por todas as orações e palavras de conforto. Aos meus tios e primos, especialmente a Marcela, por toda a torcida e apoio.

Aos meus filhos de quatro patas, Aurora e Mendel, por serem meu suporte emocional.

À Dra. Eveline Teixeira Caixeta, por ser um exemplo de superação e positividade. Por todos esses anos de orientação não somente profissional, mas também para a vida, pela paciência, oportunidade e ensinamentos transmitidos.

Ao Rodrigo Barros Rocha, por toda a orientação, conselhos e palavras de incentivo e conforto ao longo da realização desse trabalho. Sem sua ajuda não seria possível.

Ao professor Moysés Nascimento, meu coorientador, que em momento algum poupou esforços para me ajudar e tranquilizar.

Ao Bruno, meu companheiro na vida e na UFV, sem ele a realização deste trabalho não seria possível.

Ao Leandro Fumes, meu melhor amigo e companheiro de alma, por toda ajuda e incentivo.

Aos meus amigos de infância e de Viçosa, por todos esses anos de amizade e companheirismo.

Aos meus amigos do Biocafê, por toda amizade e companheirismo. Em especial a Carol Andrade por ser minha anja em todos os momentos.

Aos amigos que eu conquistei durante o doutorado e ao Crossfit Viçosa, em especial Carla e Jeferson, por todos os momentos de descontração, incentivo e torcida.

À CAPES, CNPq e FAPEMIG pela concessão da bolsa de estudos e o financiamento da pesquisa.

“Continue a nadar...”
Walt Disney

RESUMO

SILVA, Letícia de Faria, D.Sc., Universidade Federal de Viçosa, agosto de 2022. **Estrutura populacional e diversidade genética de *Coffea Canephora* detectadas por marcadores moleculares.** Orientadora: Eveline Teixeira Caixeta.

A integração de marcadores moleculares consiste em estratégia eficiente para o estudo genético e auxílio dos programas de melhoramento genético da espécie cafeeira *Coffea canephora*. Nesse contexto, os marcadores microssatélites (SSR) e, mais recentemente, os marcadores de polimorfismo único (SNP) têm sido preferencialmente escolhidos e utilizados, pois agregam informações importantes ao programa. Esses marcadores apresentam reprodutibilidade e expressão codominante, bem como permitem analisar regiões polimórficas do genoma dos diferentes cafeeiros. Nos programas de melhoramento, esses marcadores podem ser utilizados para diferentes finalidades com destaque para os estudos de diversidade genética, que auxiliam na avaliação de recursos genéticos disponíveis, seleção de genitores, identificação e caracterização de acessos de germoplasma. Assim, o presente trabalho objetivou caracterizar e identificar a estrutura populacional e diversidade genética em populações de *C. canephora* utilizando marcadores moleculares. Além disso, a eficiência dos dois marcadores, SSR e SNP, foi avaliada e comparada para essa espécie cafeeira. No capítulo 1, caracterização molecular e estudo de diversidade genética foram realizados em 96 cafeeiros cultivados na Amazônia Ocidental. Dentre esses cafeeiros estão acessos do banco de germoplasma (BAG) e híbridos do programa de melhoramento da Embrapa Rondônia, bem como clones plantados pelos produtores da região. As relações genéticas por agrupamentos hierárquicos, estrutura e análises discriminantes revelaram uma distinção entre as variedades botânicas Conilon e Robusta, havendo maior diversidade na variedade Robusta. Foi possível identificar acessos com classificações errôneas, que estavam sendo considerados Conilon ou Robusta, mas que os marcadores demonstraram ser genótipos híbridos entre as duas variedades botânicas. Grande diversidade genética foi detectada na população estudada, demonstrando seu potencial para a sustentabilidade da cafeicultura da Amazônia e para a obtenção de novas cultivares de *C. canephora*. Para o estudo de comparação da eficiência de marcadores SSR e SNP para análise da estrutura genética e diversidade populacional de cafeeiros da espécie *C. canephora* (capítulo 2), foram avaliados 165 acessos do BAG da Universidade Federal de Viçosa. Entre os acessos estão incluídos cafeeiros das variedades botânicas Conilon e Robusta, além de Híbridos interpopulacionais. O resultado demonstrou que, quando

comparado com SNP, um menor número de marcadores SSR são necessários para classificar eficientemente a população em estudo, mostrando seu potencial para análises de diversidade genética. As vantagens e desvantagens de cada tipo de marcador foi discutida. Dessa forma, os resultados obtidos nesse estudo mostraram a importância dos marcadores moleculares para estudos genéticos e para auxiliar a condução de programas de melhoramento de *C. canephora*. No presente trabalho, foram disponibilizadas informações e conhecimentos para a manutenção e uso dos recursos genéticos, fornecendo subsídios para a seleção de plantas e desenvolvimento de novas cultivares.

Palavras-chave: Melhoramento do cafeeiro. Banco de germoplasma. Marcadores SSR. Marcadores SNP.

ABSTRACT

SILVA, Leticia de Faria, D.Sc., Universidade Federal de Viçosa, August 2022. **Assessment of population structure and genetic diversity of *Coffea canephora* by molecular markers.** Adviser: Eveline Teixeira Caixeta.

The integration of molecular markers is an efficient strategy for the genetic study and support of genetic improvement programs for the coffee species *Coffea canephora*. In this context, microsatellite markers (SSR) and, more recently, single polymorphism markers (SNP) have been preferentially chosen and used, as they add important information to the program. These markers show reproducibility and codominant expression, as well as allowing the analysis of polymorphic regions of the genome of different coffee plants. In breeding programs, these markers can be used for different purposes, with emphasis on genetic diversity studies, which help in the evaluation of available genetic resources, selection of parents, identification and characterization of germplasm accessions. Thus, the present work aimed to characterize and identify the population structure and genetic diversity in populations of *C. canephora* using molecular markers. In addition, the efficiency of the two markers, SSR and SNP, was evaluated and compared for this coffee species. In chapter 1, molecular characterization and genetic diversity study were carried out in 96 coffee trees of interest to the Amazon. Among these coffee trees are accessions from the germplasm bank (BAG) and hybrids from the Embrapa Rondônia breeding program, as well as clones planted by producers in the region. Genetic relationships by hierarchical clusters, structure and discriminant analysis revealed a distinction between the botanical varieties Conilon and Robusta, with greater diversity in the Robusta variety. It was possible to identify accessions with erroneous classifications, which were being considered Conilon or Robusta, but that the markers showed to be hybrid genotypes between the two botanical varieties. Great genetic diversity was detected in the studied population, demonstrating its potential for the sustainability of coffee production in the Amazon and for obtaining new cultivars of *C. canephora*. For the study to compare the efficiency of SSR and SNP markers to analyze the genetic structure and population diversity of coffee trees of the species *C. canephora* (chapter 2), 165 accessions from the BAG of the Universidade Federal de Viçosa were used. Among the accessions, coffee plants of the botanical varieties Conilon and Robusta were included, as well as interpopulation hybrids. The result showed that, when compared to SNP, a smaller number of SSR markers are needed to efficiently classify the population under study, showing its potential for genetic diversity

analysis. The advantages and disadvantages of each type of marker were discussed. Thus, the results obtained in this study showed the importance of molecular markers for genetic studies and to assist in the conduct of breeding programs for *C. canephora*. Important knowledge concerning germplasm management, selection of genotypes for crosses, selection of the best molecular marker to be used and genetic diversity were made available.

Keywords: Coffee breeding. Germplasm collection. Molecular markers. SSR marker. SNP marker.

SUMÁRIO

1	Introdução Geral.....	12
	Referências.....	15
2	Capítulo I..... Avaliação da diversidade genética e Fingerprint de Coffea canephora do Estado de Rondônia	18
	Resumo	19
	Abstract.....	21
2.1	Introdução	22
2.2	Material e Métodos	23
2.2.1	Material vegetal.....	23
2.2.2	Extração do DNA genômico	24
2.2.3	Genotipagem com marcadores SSR.....	24
2.2.4	Análises estatísticas.....	30
2.3.1	Genotipagem com marcadores SSR.....	30
2.3.2	Diversidade genética e perfil molecular.....	32
2.4	Discussão	40
2.4.1	Genotipagem com marcadores SSR.....	40
2.4.2	Diversidade genética e perfil molecular.....	41
2.5	Conclusão.....	46
2.6	Referências.....	47
3	Capítulo II Estrutura populacional e diversidade genética do banco de germoplasma de Coffea canephora avaliado com marcadores SSR e SNP.....	51
	Resumo	52
	Abstract.....	53
3.1	Introdução	54
3.2	Material e métodos.....	55
3.2.1	Material vegetal	55
3.2.2	Extração do DNA genômico.....	59
3.2.3	Genotipagem com marcadores SSR	59
3.2.4	Genotipagem com marcadores SNP	60
3.2.5	Análises estatísticas	61
3.3	Resultados.....	61
3.3.1	Comparação entre estimativas de diversidade genética para os marcadores SSR e SNP	61
3.3.2	Descrição da estrutura populacional.....	64
3.4	Discussão	66

3.4.1	Comparação entre estimativas de diversidade genética para os marcadores SSR e SNP	66
3.4.2	Descrição da estrutura populacional.....	68
3.5	Conclusões	70

1 Introdução Geral

O café, pertencente à família Rubiaceae e ao gênero *Coffea*, possui aproximadamente 124 espécies conhecidas (DAVIS, 2011). Dentre essas espécies, *Coffea arabica* e *Coffea canephora* são as cultivadas comercialmente (BATES, 2020). O café *C. canephora* apresenta um conjunto cromossômico diploide ($2n=2x=22$), com autoincompatibilidade do tipo gametofítica, possuindo, portanto, reprodução por fecundação cruzada (HUDED *et al.*, 2020). Embora o café arábica seja suscetível a muitas pragas e doenças a qualidade de sua bebida é considerada superior em comparação ao café canéfora. No entanto as plantas de *C. canephora* além de apresentam resistência às principais doenças do cafeeiro (ANAGBOGU *et al.*, 2019), também apresentam maiores quantidades de cafeína, sólidos solúveis e oferece corpo a bebida sendo comercialmente explorada em misturas (*blends*) com *C. arabica* fato esse que impulsionou a sua produção mundial (DAVIS *et al.*, 2020).

Coffea canephora é originário da floresta tropical de várzea da África e se estende por todo território africano ocidental, passando por Uganda até o norte da Angola (MUSOLI *et al.*, 2009). A diversidade dessa espécie foi descrita em Berthaud, 1985, que a dividiu em dois grupos de acordo com seus respectivos centros de diversidade. O grupo guineano compreende os genótipos da região oeste da África e apresentam folhas menores, menor vigor e porte, frutos pequenos, tolerância a seca e são suscetíveis a ferrugem. O segundo grupo denominado de Congolês compreende os genótipos da região central da África, esse grupo pode ser dividido ainda em cinco subgrupos em consequência de seu processo de domesticação, sendo denominados SG1, SG2, B, C e mais recentemente o subgrupo UW composto por acessos selvagens de Uganda (HUDED *et al.*, 2020).

O subgrupo SG1 reúne os genótipos denominados de Conilons (Kouillous), os cafeeiros Conilon são mais tolerantes à seca e constituem as principais variedades desenvolvidas no Brasil (ALKIMIM *et al.*, 2021), O subgrupo SG2 B e C reúnem os genótipos do tipo Robusta, que são caracterizados por serem plantas mais altas, vigorosas, de folhas e frutos maiores, resistência a ferrugem e mais sensíveis a seca. Durante o processo de domesticação da espécie, o grupo guineano ficou mais restrito a seu local de origem enquanto o grupo Congolês foi disseminado para outros países e assim progressivamente melhorado nos países produtores (MEROT-L'ANTHOENE *et al.*, 2019). O processo de domesticação da espécie combinados com sua forma de reprodução e ampla distribuição geográfica (maior do gênero *Coffea*) permitiram que a espécie apresentasse uma grande variabilidade entre e dentro

das populações naturais. Essa grande variabilidade da espécie pode ser explorada nos programas de melhoramento da espécie. O sucesso de um programa de melhoramento, depende além de outros fatores, da variabilidade genética da população (FERRÃO *et al.*, 2021). A divergência genética presente em uma população pode ser quantificada pela diferença nas frequências alélicas entre um grupo de genitores cujo principal objetivo foi a identificação das combinações híbridas de maior efeito heterótico (FONSECA *et al.*, 2006). Essa quantificação pode ser estudada por meio de técnicas biométricas, análises dialélicas ou empregando marcadores moleculares (CENCI *et al.*, 2012).

Diferentes marcadores moleculares foram desenvolvidos e utilizados para a espécie *Coffea*, os métodos moleculares diferem principalmente na aplicação, no tipo, na quantidade de polimorfismo e no tempo e custo dispendido (SEMAGN *et al.*, 2006). Estudos realizados por (FERRÃO *et al.*, 2013) avaliaram a eficiência dos marcadores RAPD, AFLP e SSR em estudos genéticos com *C. canephora*, dentre esses marcadores os SSR foram considerados os mais informativos.

A utilidade desses marcadores SSR tem sido demonstrada em diferentes pesquisas em *C. canephora*, abrangendo análises de diversidade genética (FERRÃO *et al.*, 2013), caracterização de germoplasma (SOUZA *et al.*, 2013), desenvolvimento de mapas de ligação (LASHERMES *et al.*, 2001) e identificação de QTL por mapas de ligação (JOSÉ LUÍS *et al.*, 2022). Apesar desses marcadores serem de extrema utilidade pra os programas de melhoramento, a adaptação desses marcadores na genotipagem de alto rendimento ainda não foi realizada, afetando o seu custo-benefício (AKPERTEY *et al.*, 2021). Nesse contexto os marcadores moleculares de nucleotídeo único (SNP) são adequados para a genotipagem em larga escala em programas de melhoramento (ALKIMIM *et al.*, 2020), quando comparados com os marcadores SSR, os marcadores SNP são passíveis de automação de alto rendimento (KUMAR *et al.*, 2016) e com maior abundância e cobertura no genoma (ZHAO *et al.*, 2006). No entanto esses marcadores são menos polimórficos, podendo ser limitantes em análises de diversidade genética (HASSAN & HAMA-ALI, 2022). Na espécie *C. canephora* os estudos envolvendo marcadores SNP ainda são recentes e escassos, na espécie são relatados estudo de seleção genômica (ALKIMIM *et al.*, 2020), diversidade genética (FONSECA *et al.*, 2006) e identificação de genótipos (ANAGBOGU *et al.*, 2019) e recentemente (MEROT-L'ANTHOENE *et al.*, 2019) relataram o desenvolvimento de uma matriz de SNP de 8,5 k que podem ser usados para uma ampla variedade de aplicações.

Tendo em vista o exposto acima, o presente trabalho teve como objetivo caracterizar, identificar, discriminar e quantificar a diversidade genética em populações de *C. canephora*

com base em marcadores moleculares. Além disso, dois tipos de marcadores moleculares, SSR e SNP, foram comparados quanto a eficiência na análise de diversidade de cafeeiros dessa espécie.

Referências

- Akperthey A, Padi FK et al (2021) Effectiveness of Single Nucleotide Polymorphism markers in genotyping germplasm collections of *Coffea canephora* using KASP assay. **Frontiers in plant science**, 11: 612593. <https://doi.org/10.3389/fpls.2020.612593>
- Alkimim ER, Caixeta ET, Sousa TV et al (2020) Selective efficiency of genome-wide selection in *Coffea canephora* breeding. **Tree Genetics & Genomes**, 16(3):1-11. <https://doi.org/10.1007/s11295-020-01433-3>
- Alves EA, de Souza CA, Rocha RB et al (2020) Efeito da fermentação sobre qualidade da bebida do Café Robusta (*Coffea canephora*) cultivado na Amazônia Ocidental. **Revista Ifes Ciência** 6:159-170. <https://doi.org/10.36524/ric.v6i3.875>
- Anagbogu CF, Bhattacharjee R, Ilori C et al (2019) Genetic diversity and re-classification of coffee (*Coffea canephora* Pierre ex A. Froehner) from South Western Nigeria through genotyping-by-sequencing-single nucleotide polymorphism analysis. **Genetic Resources and Crop Evolution**, 66(3):685-696. <https://doi.org/10.1007/s10722-019-00744-2>
- Araújo ES, Arriel NHC, dos Santos RC et al (2019) Assessment of genetic variability in sesame accessions using SSR markers and morpho-agronomic traits. **Australian Journal of Crop Science** 13: 45-54. <https://doi.org/10.21475/ajcs.19.13.01.p1157>
- Benti T, Gebre E, Tesfaye K et al (2021) Genetic diversity among commercial arabica coffee (*Coffea arabica* L.) varieties in Ethiopia using simple sequence repeat markers. **Journal of Crop Improvement** 35:147-168. <https://doi.org/10.1080/15427528.2020.1803169>
- Cenci A, Combes MC, Lashermes, P (2012) Genome evolution in diploid and tetraploid *Coffea* species as revealed by comparative analysis of orthologous genome segments. **Plant molecular biology**, 78(1): 135-145. <https://doi.org/10.1007/s11103-011-9852-3>
- Charrier A, Berthaud, J (1985) Botanical classification of coffee. In *Coffee*. Springer, Boston, MA, pp 13-47

Combes MC, Andrzejewski S, Anthony F et al (2000) Characterization of microsatellite loci in *Coffea arabica* and related coffee species. **Molecular ecology** 9:1178-1180. <https://doi.org/10.1046/j.1365-294X.2000.00954-5.x>

COMPANHIA NACIONAL DE ABASTECIMENTO (2022) **Acompanhamento da Safra Brasileira: Café Boletim da Safra 2021 de café**. Acessado em 20 de junho 2022

Da Fonseca AFA, Sedyama T, Cruz CD et al (2006) Divergência genética em café conilon. **Pesquisa Agropecuária Brasileira** 41:599-605. <https://doi.org/10.1590/S0100-204X2006000400008>

Dangl GS, Yang J, Golino DA et al (2009) A practical method for almond cultivar identification and parental analysis using simple sequence repeat markers. **Euphytica** 168: 41–48. <https://doi.org/10.1007/s10681-008-9877-0>

Davis AP (2011) *Psilanthus manni*, the type species of *Psilanthus*, transferred to *Coffea*. **Nordic Journal of Botany** 29(4): 471-472. <https://doi.org/10.1111/j.1756-1051.2011.01113.x>

Diniz LE, Sakiyama NS, Lashermes P et al (2005) Analysis of AFLP markers associated to the Mex-1 resistance locus in Icatu progenies. **Crop Breed Appl Technol** 5(4): 387. <https://doi.org/10.12702/1984-7033.v05n04a031>

Dubberstein D, Partelli FL, Guilhen JHS et al (2020) Biometric traits as a tool for the identification and breeding of *Coffea canephora* genotypes. **Genet Mol Res**. <https://doi.org/10.4238/gmr18541>

Espindula MC, Araújo LFBD, Schmidt R et al (2020) Early induction of orthotropic shoots in *Coffea canephora*. **Revista Ceres** 67: 281-287. <https://doi.org/10.1590/0034-737X202067040005>

Ferrão LFV, Caixeta ET, Souza FDF et al (2013) Comparative study of different molecular markers for classifying and establishing genetic relationships in *Coffea canephora*. **Plant systematics and evolution**, 299(1):225-238. Disponível em: <https://doi.org/10.1007/s00606->

012-0717-2

Ferrão LFV, Ferrão RG, Ferrão MAG et al (2019) Accurate genomic prediction of *Coffea canephora* in multiple environments using whole-genome statistical models. **Heredity** 122(3): 261-275. <https://doi.org/10.1038/s41437-018-0105-y>

2 Capítulo I

AVALIAÇÃO DA DIVERSIDADE GENÉTICA E *FINGERPRINT* DE *Coffea canephora* DO ESTADO DE RONDÔNIA

Resumo

A caracterização dos recursos genéticos é fundamental para a eficiência do melhoramento genético e manejo do germoplasma de uma cultura. Essa caracterização permite o conhecimento do material genético disponível e pode ser realizada por meio de análise morfológica, bioquímica e molecular. Nesse contexto, o uso de marcadores moleculares tem sido fundamental para validação de passaportes genéticos baseados em análises de DNA, além da identificação eficientes de genótipos e rastreabilidade nos programas de melhoramento do cafeeiro. No presente trabalho, foram analisados 96 genótipos de *Coffea canephora* do Estado de Rondônia, incluindo as variedades botânicas Conilon, Robusta e Híbridos do banco de germoplasma da Empresa Brasileira de Pesquisa Agropecuária (Embrapa), além de clones de domínio público proveniente de produtores do Estado. Os cafeeiros foram avaliados com marcadores microssatélites (SSR) que permitiram discriminar todos os genótipos. Dessa forma, foram obtidos 96 perfis únicos dos cafeeiros e um conjunto de marcadores eficiente para a discriminação de cafeeiros da espécie *C. canephora*. As relações genéticas identificadas por agrupamentos hierárquicos, estrutura e análises discriminantes revelaram uma distinção entre as variedades botânicas Conilon e Robusta, havendo maior diversidade na variedade Robusta. Foi possível identificar, ainda, acessos com classificações errôneas, que estavam sendo considerados Conilon ou Robusta, mas que os marcadores demonstraram ser genótipos híbridos entre as duas variedades botânicas. Grande diversidade genética foi detectada nesse banco de germoplasma, demonstrando seu grande potencial para obtenção de novas cultivares de *C. canephora*. Por meio da análise de acessos híbridos foi possível identificar os mais similares geneticamente com uma das variedades botânicas. Esses dados são de grande importância para escolha dos materiais genéticos para comporem ou dar origem às futuras populações em melhoramento genético. As análises moleculares possibilitaram, ainda, o entendimento da genealogia de alguns acessos híbridos, que eram classificados como irmão completos ou meios-irmãos. Com base nos marcadores, observou-se que essas populações não são resultantes dos genitores inicialmente considerados, visto que possuem alelos distintos dos genitores. Esse estudo demonstrou, também, o perfil genético único de cada clone de domínio público do Estado, anteriormente desconhecido, evidenciando quais clones apresentavam maiores similaridades entre si e entre os acessos Conilon, Robusta e Híbridos do banco de germoplasma. Esses resultados destacam a utilidade de análises de marcadores moleculares para auxiliar na condução, seleção, manutenção e cruzamentos em programas de melhoramento do cafeeiro.

Palavras-chave: melhoramento genético do cafeeiro, banco de germoplasma, marcadores SSR; caracterização molecular; rastreabilidade.

Abstract

The characterization of genetic resources is fundamental for the efficiency of genetic improvement and management of the germplasm of a crop. This characterization allows the knowledge of the available genetic material and can be performed through morphological, biochemical and molecular analysis. In this context, the use of molecular markers has been fundamental for the validation of genetic passports based on DNA analysis, in addition to the efficient identification of genotypes and traceability in coffee breeding programs. In the present work, 96 genotypes of *Coffea canephora* from the State of Rondônia were analyzed, including the botanical varieties Conilon, Robusta and Híbridos from the germplasm bank of Empresa Brasileira de Pesquisa Agropecuária (Embrapa), in addition to clones in the public domain from producers in Rondônia. The coffee plants were evaluated with microsatellite markers (SSR) that allowed the discrimination of all genotypes. Thus, 96 unique profiles of coffee trees and a set of efficient markers were obtained for the discrimination of coffee trees of the species *C. canephora*. The genetic relationships identified by hierarchical groupings, structure and discriminant analysis revealed a distinction between the botanical varieties Conilon and Robusta, with a greater diversity in the Robusta variety. It was also possible to identify accessions with erroneous classifications, which were being considered Conilon or Robusta, but that the markers showed to be hybrid genotypes between the two botanical varieties. Great genetic diversity was detected in this germplasm bank, demonstrating its great potential for obtaining new cultivars of *C. canephora*. Through the analysis of hybrid accessions, it was possible to identify the most genetically similar to one of the botanical varieties. These data are of great importance for choosing genetic materials to compose or give rise to future populations in genetic improvement. Molecular analyses also made it possible to understand the genealogy of some hybrid accessions, which were classified as full or half-sibs. Based on the markers, it was observed that these populations are not the result of the parents initially considered, since they have different alleles from the parents. This study also demonstrated the unique genetic profile of each clone in the public domain in Rondônia, previously unknown, showing which clones presented greater similarities between themselves and between the Conilon, Robusta and Hybrid accessions of the germplasm bank. These results highlight the usefulness of molecular marker analysis to assist in the conduction, selection, maintenance and crosses in coffee breeding programs.

Keywords: Coffee breeding; germplasm collection; SSR markers; molecular characterization.

2.1 Introdução

Estratégias eficientes para detecção e uso da variabilidade genética em uma população são determinantes para o sucesso de um programa de melhoramento, principalmente quando se trata de espécies perenes como o café (*Coffea* spp.), cujo ciclo de vida longo necessita de maiores dispêndios de tempo e recursos (BENTI *et al.*, 2021). Nesse contexto, o cafeeiro da espécie *Coffea canephora* se destaca por ser uma planta alógama, auto incompatível que apresenta naturalmente alta diversidade genética (FERRÃO *et al.*, 2019). A espécie *C. canephora* apresenta um conjunto cromossômico diploide ($2n=2x=22$), com autoincompatibilidade do tipo gametofítica, (ALVES *et al.*, 2020), além de apresentar duas variedades botânicas, que são amplamente cultivadas e comercializadas, sendo elas: Conilon e Robusta (ALKIMIM *et al.*, 2020).

A variedade botânica Conilon possui como características um menor porte, maior tolerância a seca e suscetibilidade a pragas e doenças. Já a variedade Robusta apresenta maior porte, menor tolerância a seca e maior resistência a doenças, como a ferrugem do cafeeiro (*Hemileia vastatrix*) (ANAGBOGU *et al.*, 2019). Devido a essas características contrastantes e complementares, o desenvolvimento de híbridos entre esses dois grupos varietais tem potencial para a obtenção de novas cultivares com desempenho superior. A heterose ou vigor híbrido expresso na hibridização entre esses cafeeiros divergentes podem, portanto, serem explorados (RESENDE *et al.*, 2016; ALVES *et al.*, 2020). Assim, os programas de melhoramento da espécie dependem vigorosamente dessa variabilidade genética entre e dentro dos seus grupos varietais. ALKIMIM *et al.* (2021) demonstrou ganhos potenciais de 17% para características de interesse do cafeeiro em populações híbridas, constatando que a melhor estratégia para a condução do programa de melhoramento de *C. canephora* seria uma seleção dentro da população híbrida porque a variabilidade genética e a herdabilidade foram maximizadas na população cruzada.

Deste modo, conhecer a variabilidade genética de acessos de germoplasma e de cafeeiros cultivados e pertencentes a programas de melhoramento de *C. canephora* é de suma importância para assegurar uma maior eficiência de polinização e, conseqüentemente, obtenção de cultivares melhoradas (DUBBERSTEIN *et al.*, 2020).

Na Amazônia Ocidental, o Estado de Rondônia vem se destacando no cenário nacional como segundo maior produtor do café canéfora (CONAB, 2022) e se consolidando pelo

desenvolvimento de novas cultivares se beneficiando entre outras características pelo melhoramento dessas variedades híbridas (ESPINDULA *et al.*, 2020). A cafeicultura de Rondônia passou por intensos processos de transformações ao longo dos anos. Por exemplo, a utilização de mudas propagadas vegetativamente por meio de estaquias (clonagem) permitiu que as lavouras em termos genéticos obtivessem ganhos maiores e mais rápidos quando comparados com outras culturas perenes (ALVES *et al.*, 2020). Outra característica marcante da cafeicultura Rondoniense se encontra no fato de que os clones mais cultivados do Estado foram selecionados pelos próprios agricultores da região, sendo a Embrapa a responsável pelo levantamento, identificação e caracterização dos clones (ESPINDULA *et al.*, 2020).

A caracterização dos clones de *C. canephora* em Robusta, Conilon e Híbridos das duas variedade botânicas vem sendo realizadas, em geral, por meio de avaliações fenotípicas e em estudos de genealogia que deram origem aos clones (MORAES *et al.*, 2018). Porém, em virtude de pertencerem a mesma espécie, avaliações fenotípicas podem ser dificultadas (OLIVEIRA *et al.*, 2021). As limitações na avaliação dos indivíduos híbridos entre os dois grupos varietais podem ser ainda maiores, pois ocorre segregação das características fenotípicas típicas de cada grupo, dificultando a caracterização (ESPINDULA *et al.*, 2020). Assim, o uso de marcadores moleculares que podem detectar alterações a nível de DNA são ferramentas genéticas altamente confiáveis para complementar a caracterização fenotípica das plantas (DANGL *et al.*, 2009). Os marcadores moleculares SSR (repetição de sequência simples, do inglês *Simple Sequence Repeats*) tem sido amplamente utilizados na análise da diversidade genética do café, estrutura populacional e formação de grupos heteróticos (FERRÃO *et al.*, 2019). Esses marcadores são utilizados por detectarem, com eficiência, o polimorfismo no DNA, apresentarem reprodutibilidade e abundância nos genomas, além de serem marcadores co-dominantes (TEKLU *et al.*, 2022)

Diante do exposto, objetivou-se analisar, discriminar e caracterizar os genótipos da espécie *C. canephora* que são utilizados por produtores do Estado de Rondônia, bem como clones melhorados do programa de melhoramento da Embrapa Rondônia.

2.2 Material e Métodos

2.2.1 Material vegetal

Foram coletadas amostras foliares de 96 clones de *C. canephora* cultivados no Estado de Rondônia (Tabela 1). Entre os clones estudados estão as 10 cultivares lançadas pela Embrapa (BRS) e cafeeiros das duas variedades botânicas, Conilon (12 genótipos) e Robusta

(21 genótipos), que pertencem ao Banco de Germoplasma da Embrapa Rondônia. Os Robustas foram originados da coleção do Instituto Agronômico de Campinas (IAC), sendo derivados de sementes das progênes IAC 1655, IAC 2258-1, IAC 2259, IAC Col-5 e IAC Col-10 e da cultivar Apoatã IAC 2258. Os Conilons foram introduzidos do Instituto Capixaba de Pesquisa, Assistência Técnica e Extensão Rural (Incaper). Foram incluídos na análise, cafeeiros híbridos de polinização aberta (oito genótipos coletados em lavoura de Robusta) e híbridos controlados (sete genótipos híbridos entre Robusta e Conilon) do programa de melhoramento da Embrapa, além de 38 clones de produtores do Estado de Rondônia, os quais são de domínio público e amplamente cultivados no Estado.

2.2.2 Extração do DNA genômico

Para cada genótipo selecionado, folhas saudáveis, totalmente expandidas, jovens e de coloração verde clara foram coletadas, liofilizadas e maceradas. O DNA genômico foi extraído pelo método proposto por DINIZ *et al.* (2005). A qualidade e quantidade do DNA foram avaliadas usando o espectrofotômetro NanoDrop 2000-Thermo Scientific. As amostras foram padronizadas a 25 ng. μL^{-1} e armazenadas a $-20\text{ }^{\circ}\text{C}$.

2.2.3 Genotipagem com marcadores SSR

Foram utilizados *primers* microssatélites (SSR) (Tabela 2), sendo parte disponibilizada na literatura, alguns desenvolvidos a partir de dados do Projeto Brasileiro do Genoma Café e outros de biblioteca genômica enriquecida, estes dois últimos desenvolvidos no Laboratório BioCafé/UFV (COMBES *et al.*, 2000; LASHERMES *et al.*, 2001; KIGUONGO *et al.*, 2014; DAVEY *et al.*, 2015). Inicialmente esses *primers* foram desenvolvidos para a espécie *C. arábica*, assim a fim de conferir o alinhamento, posição e a taxa de *gap* dos *primers* em estudos, todos os *primers* selecionados foram testados pela ferramenta de PCR eletrônico (e-PCR) (ROTMISTROVSKY *et al.*, 2004).

A amplificação dos *primers* foi realizada da seguinte forma: 50ng de DNA, 0,6 unidades de Taq DNA polimerase (Invitrogen), tampão enzimático 1x, 1mM de MgCl_2 , 150 μM de cada dNTP e 0,1 μM de cada *primer*, completando o volume total de 20 μl com água milli-Q estéril. As reações foram realizadas nos termocicladores PTC200 (MJ Research) e Veriti (Applied Biosystems). Após a desnaturação inicial a 94°C por dois minutos, foram

realizados 13 ciclos de PCR de desnaturação a 94°C por 30 segundos, anelamento com temperatura decrescente de 1°C para cada ciclo (67 a 55°C) por 30 segundos, e extensão a 72°C por 30 segundos. Em seguida foram realizados outros 30 ciclos composto por desnaturação a 94°C, anelamento a 55°C e extensão a 72°C por 30 segundos em cada estágio. A extensão final foi realizada a 72°C por trinta minutos. Os produtos resultantes da reação de PCR foram visualizados por eletroforese capilar, utilizando o analisador de DNA ABI 3130X da Applied Biosystem (Applied Biosystems, Foster City, CA, EUA). Os fragmentos amplificados pelos *primers* SSR (padrão codominante) foram comparados com um padrão GeneScan 500ROX (Applied Biosystems, Foster City, CA, EUA) e posteriormente codificados de acordo com o tamanho dos fragmentos gerados. Dessa forma, os dados foram analisados no programa GeneMapper versão 4.0, da Applied Biosystem (Applied Biosystems, Foster City, CA, EUA), e as codificações foram feitas a partir do tamanho apresentado por cada alelo (pico).

Tabela 1. Genótipos de cafeeiros analisados, pertencentes ao banco de germoplasma e programa de melhoramento da Embrapa Rondônia, bem como clones provenientes de produtores do Estado de Rondônia (clones de domínio público).

Amostra	Identificação	Genealogia	Descrição
1	BRS 2299		Cultivar registrada
2	BRS 1216	Emcapa03 x Robusta1675	Cultivar registrada
3	BRS 3213	Emcapa03 x Robusta2258	Cultivar registrada
4	BRS 3210	Emcapa03 x Robusta2258	Cultivar registrada
5	BRS 3220	Emcapa03 x Robusta1675	Cultivar registrada
6	BRS 2314	Emcapa03 x Robusta640	Cultivar registrada
7	BRS 3193	Robusta x Conilon	Cultivar registrada
8	BRS 2336	Robusta x Conilon	Cultivar registrada
9	BRS 3137	Robusta x Conilon	Cultivar registrada
10	BRS 2357	Robusta x Conilon	Cultivar registrada
11	H1-BAG	Robusta (ANX998)	Híbrido de polinização aberta
12	H2-BAG	Robusta (ANX 1059)	Híbrido de polinização aberta
13	H3-BAG	Robusta (ANX 1044)	Híbrido de polinização aberta
14	H4-BAG	Robusta (T13R1C6P1)	Híbrido de polinização aberta
15	H5-BAG	Robusta (T14R2C4PU)	Híbrido de polinização aberta
16	H6-BAG	Robusta (T8R1C4P2)	Híbrido de polinização aberta
17	H7-BAG	Robusta (T16R1C8P1)	Híbrido de polinização aberta
18	H8-BAG	Robusta (B3T8P5)	Híbrido de polinização direcionada
19	Clone 08	?	Clone de domínio público (produtor)
20	H9-BAG	Robusta x Conilon (B3T7P6)	Híbrido de polinização direcionada
21	H10-BAG	Robusta x Conilon (B1T9P1)	Híbrido de polinização direcionada
22	H11-BAG	Robusta x Conilon (B4T8P8)	Híbrido de polinização direcionada
23	H12-BAG	?	Clone de domínio público (produtor)
24	H13-BAG	Robusta x Conilon (B3T7P1)	Híbrido de polinização direcionada
25	Clone 25	?	Clone de domínio público (produtor)
26	H14-BAG	Robusta x Conilon (B3T8P3)	Híbrido de polinização direcionada
27	H15-BAG	Robusta x Conilon (B3T8P5)	Híbrido de polinização direcionada
28	Clone N13	?	Clone de domínio público (produtor)
29	Clone N12	?	Clone de domínio público (produtor)
30	Clone N8(G8)	?	Clone de domínio público (produtor)
31	Clone N02	?	Clone de domínio público (produtor)
32	Clone N01	?	Clone de domínio público (produtor)
33	Clone AS2	?	Clone de domínio público (produtor)
34	H16-BAG	Robusta (T8R1P4PU)	Híbrido de polinização aberta (Robusta)
35	Clone N16	?	Clone de domínio público (produtor)
36	Clone AS1	?	Clone de domínio público (produtor)
37	Clone AS5	?	Clone de domínio público (produtor)
38	Clone AS6	?	Clone de domínio público (produtor)
39	Clone AS7	?	Clone de domínio público (produtor)
40	Clone AS10	?	Clone de domínio público (produtor)

Tabela 1. continuação...

Amostra	Identificação	Clone	Descrição
41	R1-BAG	Robusta (R128)	Robusta derivado de Progênie do IAC
42	R2-BAG	Robusta (R217)	Robusta derivado de Progênie do IAC
43	R3-BAG	Robusta (R124)	Robusta derivado de Progênie do IAC
44	R4-BAG	Robusta (R126)	Robusta derivado de Progênie do IAC
45	R5-BAG	Robusta (R32)	Robusta derivado de Progênie do IAC
46	R6-BAG	Robusta (R36)	Robusta derivado de Progênie do IAC
47	R7-BAG	Robusta (R272)	Robusta derivado de Progênie do IAC
48	R8-BAG	Robusta (R243)	Robusta derivado de Progênie do IAC
49	R9-BAG	Robusta (R65)	Robusta derivado de Progênie do IAC
50	R10-BAG	Robusta (R66)	Robusta derivado de Progênie do IAC
51	R11-BAG	Robusta (R63)	Robusta derivado de Progênie do IAC
52	R12-BAG	Robusta (R224)	Robusta derivado de Progênie do IAC
53	R13-BAG	Robusta (R211)	Robusta derivado de Progênie do IAC
54	R14-BAG	Robusta (AP1)	Robusta derivado da cultivar Apoatã
55	R15-BAG	Robusta (AP2)	Robusta derivado da cultivar Apoatã
56	R16-BAG	Robusta (AP3)	Robusta derivado da cultivar Apoatã
57	R17-BAG	Robusta (AP4)	Robusta derivado da cultivar Apoatã
58	R18-BAG	Robusta (AP5)	Robusta derivado da cultivar Apoatã
59	R19-BAG	Robusta (AP6)	Robusta derivado da cultivar Apoatã
60	R20-BAG	Robusta (AP7)	Robusta derivado da cultivar Apoatã
61	R21-BAG	Robusta (AP8)	Robusta derivado da cultivar Apoatã
62	AS12	?	Clone de domínio público (produtor)
63	Clone 01	?	Clone de domínio público (produtor)
64	Clone 04	?	Clone de domínio público (produtor)
65	Clone 06	?	Clone de domínio público (produtor)
66	Clone 07	?	Clone de domínio público (produtor)
67	LB010	?	Clone de domínio público (produtor)
68	LB015	?	Clone de domínio público (produtor)
69	LB80	?	Clone de domínio público (produtor)
70	R22	?	Clone de domínio público (produtor)
71	R152	?	Clone de domínio público (produtor)
72	P42	?	Clone de domínio público (produtor)
73	Clone 106	?	Clone de domínio público (produtor)
74	Clone 03	?	Clone de domínio público (produtor)
75	Clone 05	?	Clone de domínio público (produtor)
76	Clone 80	?	Clone de domínio público (produtor)
77	Clone 41	?	Clone de domínio público (produtor)
78	Clone 21	?	Clone de domínio público (produtor)
79	Clone 31	?	Clone de domínio público (produtor)
80	Clone P50	?	Clone de domínio público (produtor)

Tabela 1. Continuação...

Amostra	Identificação	Clone	Descrição
81	Clone BG180	?	Clone de domínio público
82	Clone AS3	?	Clone de domínio público
83	Clone L1	?	Clone de domínio público
84	H17-BAG	Robusta (ANX1083)	Híbrido de polinização aberta
85	C1-BAG	Conilon (C154)	Conilon
86	C2-BAG	Conilon (C566)	Conilon
87	C3-BAG	Conilon (C795)	Conilon
88	C4-BAG	Conilon (C1089)	Conilon
89	C5-BAG	Conilon (C46)	Conilon
90	C6-BAG	Conilon (C201)	Conilon
91	C7-BAG	Conilon (C556)	Conilon
92	C8-BAG	Conilon (C796)	Conilon
93	C9-BAG	Conilon (C854)	Conilon
94	C10-BAG	Conilon (C1048)	Conilon
95	C11-BAG	Conilon (C530)	Conilon
96	C12-BAG	Conilon (C69)	Conilon

Tabela 2. *Primers* SSR usados para genotipagem dos 96 cafeeiros de *Coffea canephora* do Estado de Rondônia.

<i>Primer</i>	<i>Tam.</i>	<i>Forward</i>	<i>Reverse</i>
SSR 07 ^a	215	TGACATAGGGGGCTAAATTG	TTAATGGTGACGCTTTGATG
SSR 16 ^b	140-146	ACCCGAAAGAAAGAACCAAG	CCACACAACCTCTCCTCATTC
SSR 34 ^c	294	GGAGACGCAGGTGGTAGAAG	TCGAGAAGTCTTGGGGTGTT
SSR 35 ^c	158	CTGGCATTAGAAAGCACCTTG	GCTTGGCTCACTGTAGGACTG
SSR 119 ^c	190	TTGCCATCATCGTTCATTCT	GCATAGTGTCGGTTGTGTTGTT
SSR 122 ^c	237	CGTCTCGTTTCACGCTCTCT	GATCTGCATGTACTGGTGCTTC
CaEST-002 ^d	184	GAAGGGACAAAGACGCCTAA	CGACAGATGCAGGAATAAACTG
CaEST-029 ^e	199	AGGAGATGCCTGTGACGAAC	GGACGGAAAGATTCTGGCTTT
CaEST-071 ^e	155	ATGGAGAGGAAGACGCAACA	CCTTATTGAAGACGCCCAA
CaEST-103 ^f	100	CGCCCGAAGATCAAACAA	TTATATCCCGCGGCAAGTCC
Sat 235 ^g	222	TCGTTCTGTCATTAATCGTCAA	GCAAATCATGAAAATAGTTGGTG
Sat 207 ^g	90	GAAGCCGTTTCAAGCC	CAATCTCTTCCGATGCTCT

Forward: sequência de nucleotídeos na direção 3' para 5', *Reverse*: sequência de nucleotídeos na direção 5' para 3'.
Referências: ^a ROVELLI *et al.*, 2000; ^b COMBES *et al.*, 2000; ^c PONCET *et al.*, 2004; ^d SOUZA *et al.*, 2013; ^e MISSIO *et al.*, 2010; ^f FERRÃO, *et al.*, 2013; ^g KIGUONGO *et al.*, 2014.

2.2.4 Análises estatísticas

As análises de diversidade genética foram realizadas usando o software R v. 4.1.3 (R Development Core and Team, 2022). As informações dos alelos gerados pelos marcadores SSR foram convertidos em um objeto da classe *genind* usando o pacote “adegenet” (JOMBART & AHMED, 2011). O número de alelos por loco (N_A), heterozigotidade observada (H_O), heterozigotidade esperada (H_E), e estatísticas F (Coeficiente de fixação de Wright's (F_{st}) e Coeficiente de parentesco (F_{is})) de cada marcador SSR foram determinados para toda a população e para cada grupo. A matriz combinação F_{ST} entre grupos foi calculada através do pacote “hierfstat” (GOUDET, 2005).

As relações genéticas entre os acessos foram determinadas usando o método de agrupamento UPGMA (Método de grupos não ponderado usando médias aritméticas) e usando a matriz de dissimilaridade genética calculada pela distância de Nei (NEI & LI, 1979). O *software* Figtree v. 1.4.3 foi utilizado para formar o dendrograma (RAMBAUT, 2016). Para a análise da estrutura populacional dos acessos de cafeeiro utilizou-se a análise discriminante de componentes principais (DAPC), um método multivariado projetado para identificar e descrever agrupamentos de indivíduos geneticamente relacionados. Essa metodologia permite extrair informações avançadas de dados genéticos, fornecendo atribuições a grupos e contribuição de alelos individuais para a estruturação da população. Essa metodologia utiliza o *software* STRUCTURE como *benchmark* (JOMBART *et al.*, 2010). Essa análise permite contabilizar a porcentagem de pertencimento de cada genótipo em sua respectiva variedade botânica de acordo com sua ancestralidade e com os marcadores SSR analisados. Ao final, foi realizado a Análise de Variância Molecular (AMOVA) usando o pacote “pegas” para calcular os componentes de variância entre grupos (PARADIS, 2010).

2.2 Resultados

2.3.1 Genotipagem com marcadores SSR

Dez pares de *primers* SSR foram selecionados para avaliar a diversidade genética e a estrutura populacional. Por meio do e-PCR, observou-se que esses marcadores, amplificaram regiões dos cromossomos 0, 1, 2, 6, 7, 9 e 11, sendo, portanto, bem distribuídos no genoma desses cafeeiros.

Um total de 62 alelos foram amplificados utilizando esses 12 marcadores SSR nos 96 cafeeiros analisados. Foi observada média de 5,16 bandas por *primer*, variando de 2 a 6 bandas. A porcentagem de polimorfismo dos *primers* foi de 43% e a taxa de dados perdidos foi de 4,89%. O número de alelos por loco (NA) variou de dois (*primers* CaEST-071, Sat 235, SSR 119) a seis (SSR 34). O menor valor de heterozigoidade observada (Ho) foi de 0,00 do *primer* SAT 235 e o maior valor de 0,78 do *primer* SSR 35, e com média entre todos os *primers* de 0,44. A heterozigoidade esperada variou de 0,42 (Sat 235) a 0,74 (SSR 07 e SSR 122), com média 0,64. Analisando as estatísticas F, há uma grande variação no F_{IS} , variando de -0,12 (SSR 35) a 1,00 (Sat 235), enquanto o F_{ST} variou de -0,016 (CaEST-071) a 0,023 (SSR 112) (Tabela 3). A matriz do índice geral de fixação F_{st} variou de 0,002 para Conilon e Robusta a 0,024 para Conilon e Híbridos (Tabela 4). A análise de variância molecular (AMOVA) (Tabela 5) demonstrou uma alta variabilidade apresentada pelos acessos em estudo, sendo 80% da variação devido a variabilidade presente entre os acessos.

Tabela 3. Número de alelos (Na), Heterozigoidade observada (Ho), Heterozigoidade esperada (He), Coeficiente de fixação de Wright's (F_{st}) e Coeficiente de parentesco (F_{is}) para os dez *primers* analisados em 96 cafeeiros da espécie *C. canephora*.

Loco	Na	Ho	He	F_{st}	F_{is}
SSR 07	6	0,34	0,74	0,002908	0,536298
SSR 16	4	0,09	0,53	0,004119	0,830977
SSR 35	6	0,78	0,68	-0,012160	-0,12609
SSR 119	3	0,53	0,55	0,002792	0,042796
SSR 112	6	0,62	0,74	0,023522	0,125443
CaEST-029	6	0,32	0,65	0,004089	0,520444
CaEST-071	3	0,49	0,60	-0,01629	0,20438
CaEST-103	6	0,53	0,69	0,00153	0,235909
SAT 235	3	0	0,42	0,00067	1,000000
SAT 207	6	0,77	0,72	-0,0116	-0,05231

Tabela 4. Matriz de combinação F_{ST} entre os grupos (Híbrido, Conilon, Robusta e Clones de produtores) baseados nos fenótipos da Embrapa

	Híbrido	Conilon	Robusta	Clones produtores
Híbrido				
Conilon	-0,0243			
Robusta	-0,0034	-0,002		
Clones	-0,0019	-0,0142	-0,0047	

Tabela 5. Análise de variância molecular (AMOVA) dos 96 acessos alocados em quatro grupos.

Fonte de variação	GL	SQ	QM	P. valor	Phi	% de variância
Dentro	3	6,54	2,18	0,033	0,004	80%
Entre	92	182,01	1,97			
Total	95	188,56	1,98			

GL: graus de liberdade, SQ: soma dos quadrados, QM: quadrado médio

2.3.2 Diversidade genética e perfil molecular

Os 10 marcadores moleculares utilizados permitiram discriminar todos os cafeeiros analisados, obtendo-se, portanto, 96 perfis moleculares distintos. Cada cafeeiro foi codificado de acordo com o tamanho (banda) apresentado por cada *primer* analisado (Tabela S1). O padrão molecular de cada genótipo pode ser utilizado para *fingerprinting* desses cafeeiros. Além disso, nesse estudo, a caracterização molecular foi usada para conhecer a diversidade genética dos cafeeiros que estão sendo mantidos e plantados no Estado de Rondônia.

Inicialmente foi realizado o agrupamento dos genótipos do banco ativo de germoplasma (BAG) da Embrapa, que são sabidamente pertencentes às variedades botânicas: Conilon e Robusta. Foram considerados nessa análise 12 cafeeiros Conilon, C1-BAG a C12-BAG, e 21 cafeeiros Robusta, R1-BAG a R21-BAG, introduzidos do IAC (Tabela 1). Por meio do dendrograma UPGMA (Figura 1) observou-se a formação de oito grupos. O agrupamento dos acessos foi consistente com a separação dos genótipos em duas populações diferentes, das variedades botânicas Conilon e Robusta

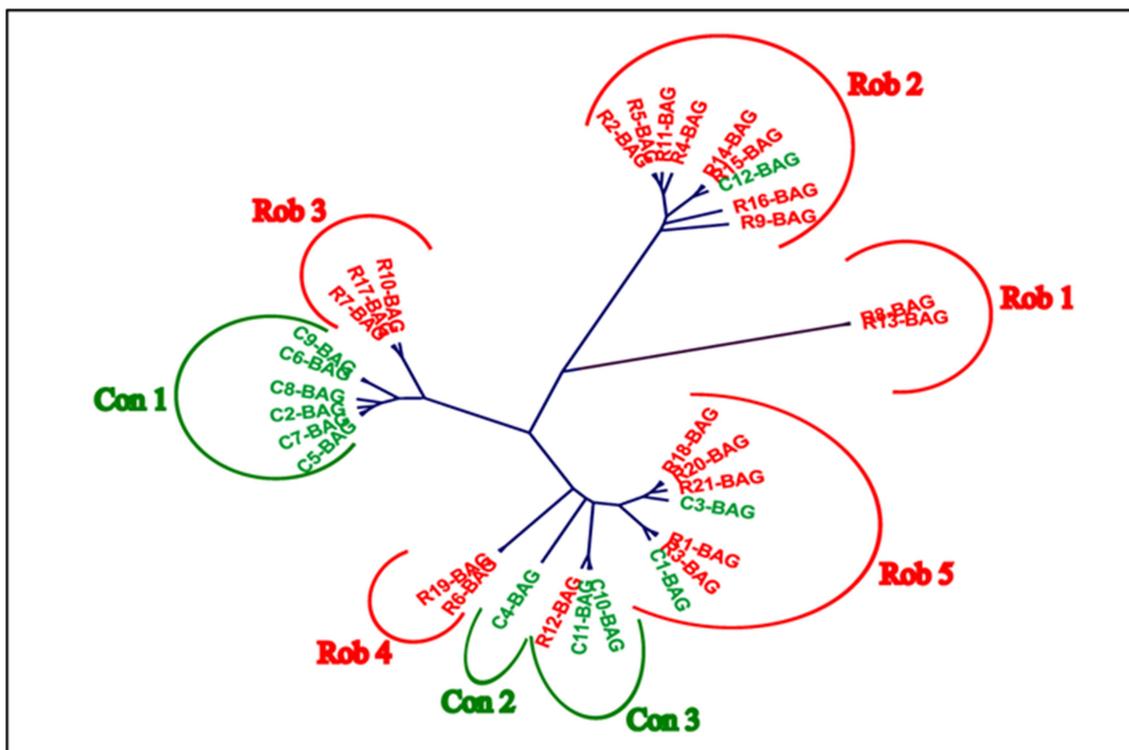


Figura 1. Dendrograma UPGMA dos acessos Conilon e Robusta de *C. canephora* do Estado de Rondônia. Clados em vermelho e verde representam as variedades botânicas: Conilon e Robusta. Em vermelho estão representados os cluster Robusta formados (Rob1, Rob2, Rob3, Rob4 e Rob 5) e em verde os cluster Conilon (Con1, Con2 e Con3).

Os genótipos Robusta foram agrupados em três grandes grupos que se caracterizaram pela presença apenas de genótipos da variedade botânica Robusta (Rob 1, Rob 3 e Rob 4) e em outros três grupos (Rob 2, Rob 5 e Rob 6) onde apesar da maior sobreposição com genótipos da variedade botânica Robusta, também se agruparam genótipos de Conilon.. Pode-se destacar ainda os genótipos R-8 BAG e R-13 BAG (Rob 1) apresentaram maior divergência em relação aos demais indivíduos. Por sua vez, os genótipos de Conilon apresentaram uma separação menos coesa que os genótipos Robusta, sendo agrupados em um grande grupo (Con 1). O genótipo C4-BAG foi alocado sozinho no grupo Con2 e o restante dos genótipos foi alocado nos grupos Rob 2, Rob 5 e Rob 6 em conjunto com os cafeeiros Robusta.

Ao se observar a Figura 2, é possível analisar a probabilidade de pertencimento de cada genótipo na sua respectiva classificação realizada com base nos fenótipos (Conilon ou Robusta). Dentre os cafeeiros da variedade botânica Conilon, C1-BAG e C3-BAG apresentaram a maior porcentagem de genoma Robusta e essa tendência também foi observada por meio da estatística de métodos de hierárquicos (Figura 1), onde no

dendrograma esse genótipo se encontra agrupado com grande similaridade com os cafeeiros Robusta R1-BAG e R3-BAG



Figura 2. Gráfico de barras da análise bayesiana de pertencimento utilizada para estudo de diversidade na espécie *C. canephora*. Os 33 genótipos estão representados na mesma ordem da Tabela 1 e foram divididos em dois grupos ($k=2$) de acordo com as variedades botânicas Conilon (Verde) e Robusta (Vermelho). O eixo x representa os cafeeiros e o eixo y a probabilidade de pertencimento em seu respectivo grupo.

Posteriormente, foi realizada a análise dos 58 cafeeiros que pertencem ao banco de germoplasma e programa de melhoramento da Embrapa. Portanto, nessa análise, foram incluídos as cultivares (clones) registradas, os cafeeiros das variedades botânicas Conilon e Robusta, respectivamente, bem como os híbridos de polinização aberta e direcionada (Tabela 1). O dendrograma (Figura 3) mostra a formação de oito grupos, sendo que nos grupos I, II e VII foram alocados cafeeiros das duas variedades botânicas, cultivares registradas (Robusta x Conilon) e híbridos de polinização aberta (Robusta) e direcionada (Robusta x Conilon).

Pela análise molecular, destaca-se ainda que o híbrido de polinização aberta H7-BAG não ficou alocado com nenhum outro genótipo (Figura 3), demonstrando sua maior diversidade entre os híbridos em estudo. Observou-se, ainda, que os clones correspondentes às cultivares registradas foram alocadas em grupos diferentes, sendo BRS 1216 e BRS 2314 no grupo I, o BRS 3213 no grupo II, o BRS 3210 no grupo V. Esses clones além de serem divergentes geneticamente, possuem grupos de compatibilidade distintos, permitindo assim seus cruzamentos para formação novas populações para o melhoramento.

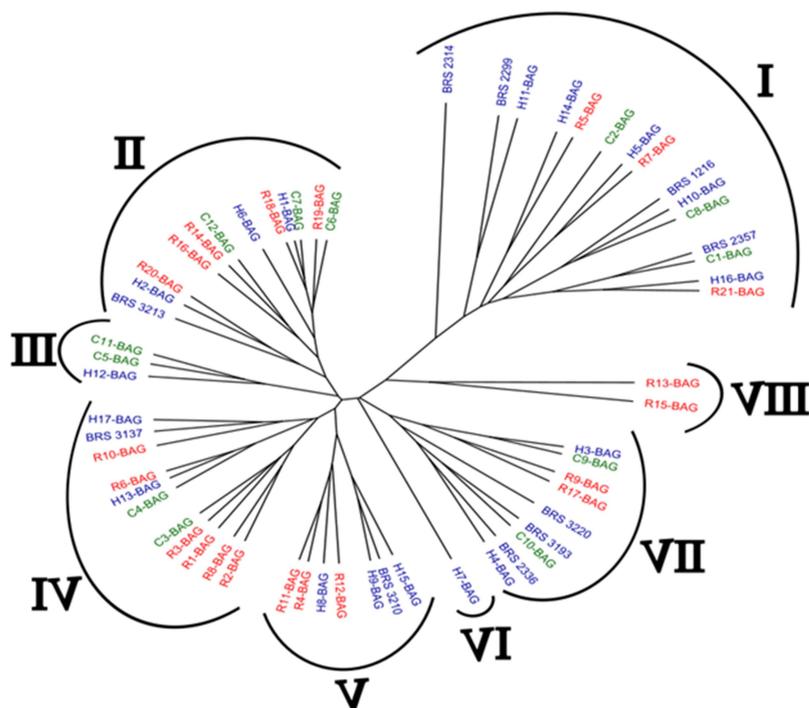


Figura 3. Dendrograma UPGMA dos acessos Conilon, Robusta e Híbridos do BAG de *C. canephora* do Estado de Rondônia. Clados em preto (I, II, III, IV, V, VI, VII e VIII) representam os *cluster* formados. Os genótipos Conilon (verde), Robusta (Vermelho) e Híbridos (Azul) estão representados pelas respectivas cores.

No presente estudo, clones de domínio público e que são cultivados pelos produtores do Estado de Rondônia foram analisados e comparados com os cafeeiros do BAG da Embrapa. Em virtude do modo que a cafeicultura de Rondônia se desenvolveu, os clones de produtores não apresentam passaportes genéticos confirmados, podendo ser materiais distintos dos cafeeiros produzidos e mantidos no BAG da Embrapa. Além disso, não se conhece a variedade botânica que pertencem e nem mesmo se são distintos entre si. Esses clones de produtores foram analisados com os cafeeiros sabidamente Conilon e Robusta do BAG (controles). A Figura 4 revela a formação de sete grupos, com grande variabilidade e formação de subgrupos. Todos os cafeeiros avaliados são geneticamente distintos, não sendo observado, portanto, um mesmo genótipo sendo cultivado por diferentes produtores. No dendrograma, observa-se que os clones dos produtores foram alocados em todos os grupos e os dados permitem identificar cafeeiros geneticamente mais próximos de Conilon e outros mais próximos de Robusta. Somente o grupo IV alocou apenas clones de produtores, indicando que esses dois genótipos se divergem dos demais clones e das variedades botânicas do BAG da Embrapa.

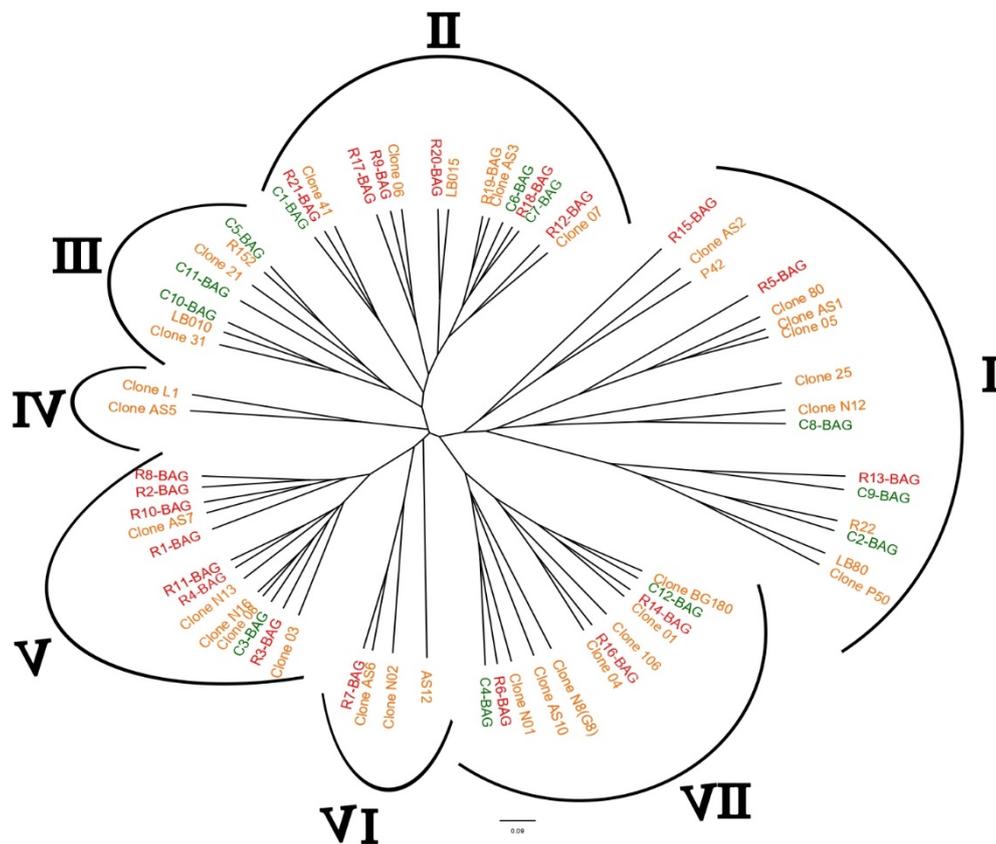


Figura 4. Dendrograma UPGMA dos clones de produtores de *C. canephora* do Estado de Rondônia. Clados em preto (I, II, III, IV, V e VI) representam os *cluster* formados. Os genótipos Conilon (verde), Robusta (Vermelho) e clones de produtores (Laranja) estão representados pelas respectivas cores.

Na análise conjunta de todos os 96 genótipos, o dendrograma UPGMA (Figura 5) mostrou a formação de quatro grupos subdivididos em subgrupos menores suportados por um alto valor de *bootstrap* (100). No grupo I ficaram alocados 47,6% dos clones de produtores, 23,8% genótipos híbridos, 19,0% de genótipos Robusta e 9,5% de genótipos Conilon. O grupo II alocou 32,2% de genótipos Robusta, 9,7% de genótipos Conilon, 25,8% de genótipos híbridos e 32,2% de clones. O grupo III alocou 11,1% de genótipos Robusta, 33,3% de genótipos híbridos, 22,2% de Conilon e 33,3% de clones. Já o grupo IV foi formado por 15,4% de genótipos Robusta, 30,7% de genótipos híbridos, 11,5% de Conilon e 38,4% de clones de domínio público. A Figura 6 mostra as porcentagens de introgressão das variedades botânicas Conilon e Robusta em cada um dos 96 acessos analisados.

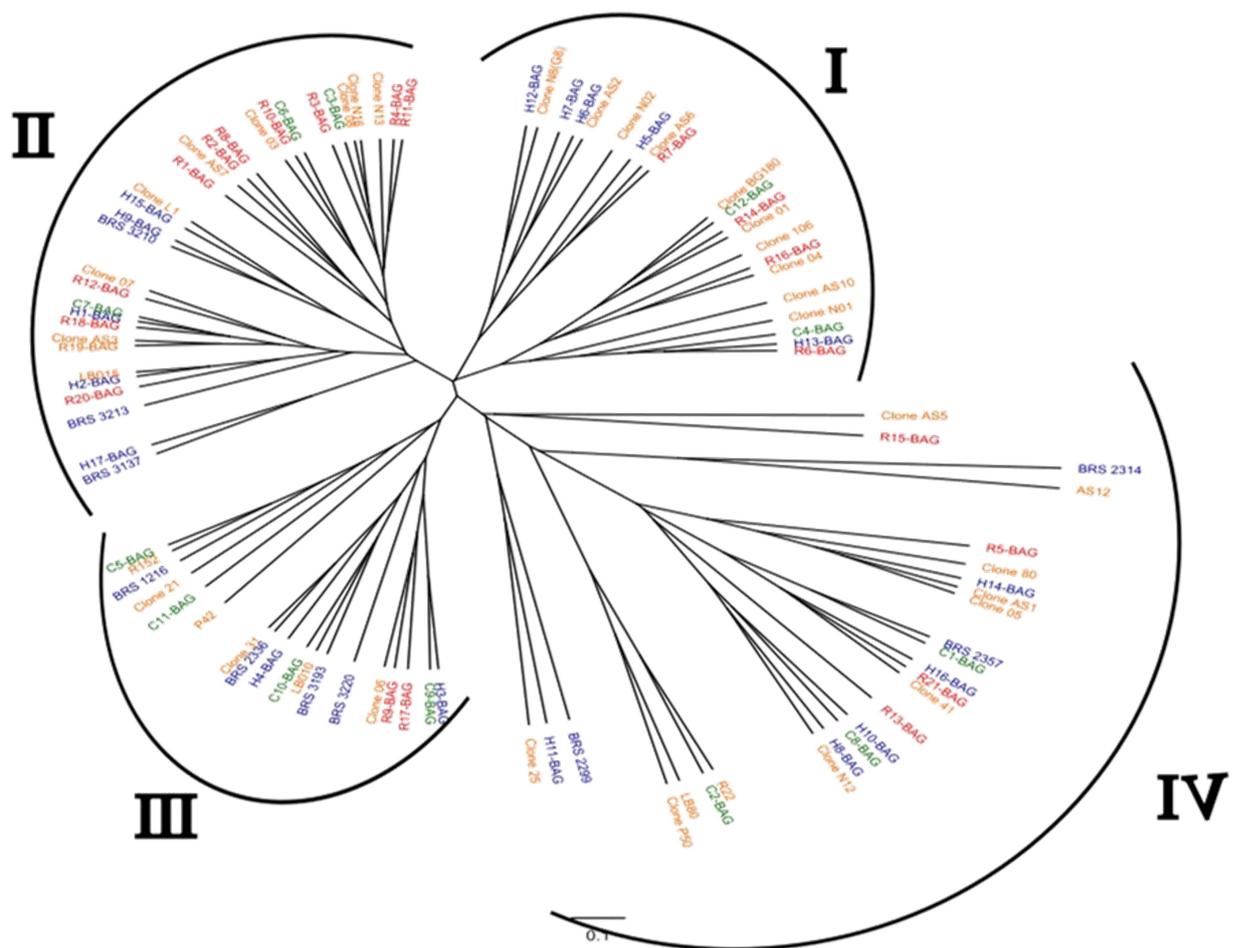


Figura 5. Dendrograma UPGMA dos 96 acessos de *C.canephora* do Estado de Rondônia. Clados em preto (I, II, III e IV) representam os *cluster* formados. Os genótipos Conilon (verde), Robusta (Vermelho), Híbridos (Azul) e clones de produtores (Laranja) estão representados pelas respectivas cores.

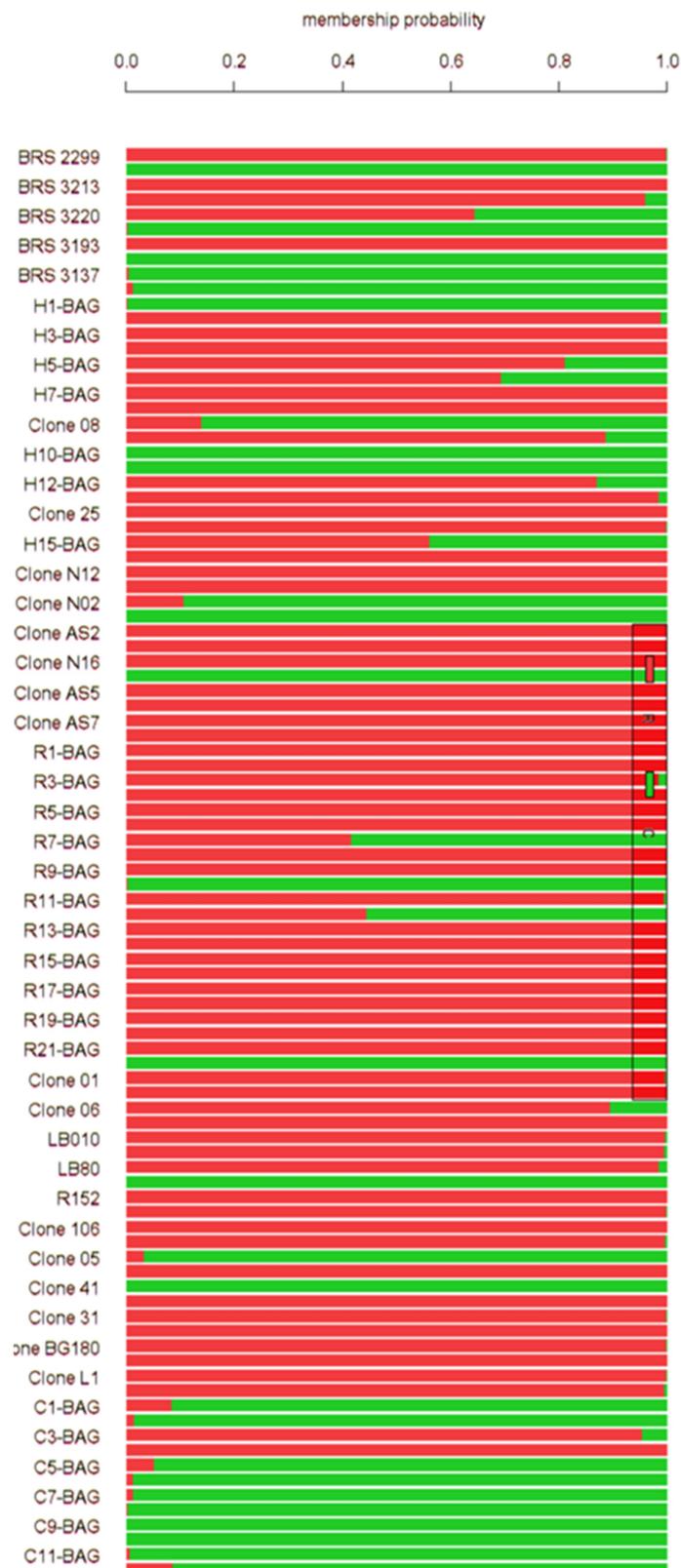


Figura 6. Gráfico de barras da análise bayesiana de pertencimento utilizada para estudo de diversidade na espécie *C.canephora*. Os 96 genótipos estão representados na mesma ordem da Tabela 1 e foram divididos em dois grupos ($k=2$) de acordo com as variedades botânicas Conilon (Verde) e Robusta (Vermelho). O eixo x representa os cafeeiros e o eixo y a probabilidade de pertencimento em seu respectivo grupo.

A fim de facilitar o entendimento sobre a estrutura populacional dos cafeeiros em estudo, as análises foram realizadas considerando grupos baseados nas avaliações fenotípicas da Embrapa, sendo eles Conilon, Robusta, Híbridos e Clones de produtores. Para validar a diferenciação genética entre os grupos de cafeeiros, foram determinados os valores de F_{ST} baseados na distância genética de Nei entre os grupos (Tabela 4). Os valores globais de F_{ST} pareados variaram de 0,002 a 0,014. Esses valores sugeriram uma diferenciação moderada entre os grupos Conilon e Robusta e a maior diferenciação observada foi entre os grupos de cafeeiros de produtores e Conilon.

A estrutura genética obtida pelo DAPC (Figura 7), empregada para confrontar a classificação dos grupos formados pelas análises fenotípicas (Conilon, Robusta, Híbridos e Clones de produtores) com a genotipagem realizada pelo marcadores SSR, demonstrou uma proporção de acerto geral dos grupos de 0,7708. As proporções individuais de acerto na classificação dos indivíduos nos grupos foi de 0,7368 para os híbridos, 0,6923 para Conilon, 0,7692 para Robusta e 0,8157 para os clones dos produtores. O nível ideal de componentes principais calculado pelo *a-score* foi de 42. O cenário da distribuição dos grupos pode ser observado na Figura 7, em que diferenciações entre os grupos foram discriminados por duas funções. As funções discriminantes foram suficientes para discriminar os quatro grupos, sendo que pela primeira função, o grupo clone de produtores e Robusta se sobrepuseram. Com base na segunda função, os híbridos sobrepuseram com o Robusta, enquanto os clones de produtores sobrepuseram com os Conilon.

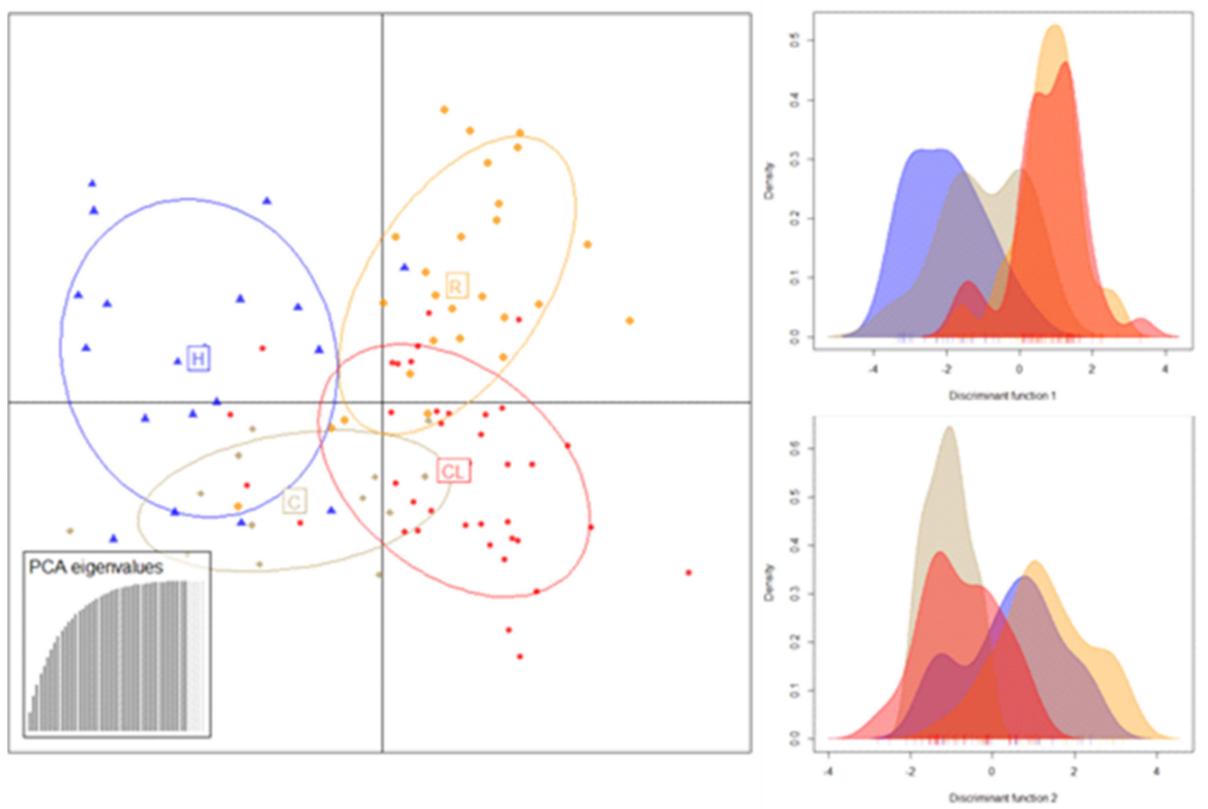


Figura 7. Scatterplot do DAPC dos 96 genótipos junto com as funções discriminantes empregada para confrontar a classificação dos grupos formados pelas análises fenotípicas (Conilon, Robusta, Híbridos e Clones de produtores).

2.3 Discussão

2.4.1 Genotipagem com marcadores SSR

A escolha dos *primers* se mostrou eficiente para a discriminação e caracterização da população de *C. canephora* em estudo. Com base no PCR eletrônico, foi verificado que os marcadores se encontram distribuídos em diferentes cromossomos, garantindo uma boa cobertura do genoma em regiões distribuídas nos cromossomos 0, 1, 2, 6, 7, 9 e 11. Na espécie *C. canephora*, o cromossomo 0 se refere a um conjunto de *Scaffolds* não ordenados que não foram agrupados em nenhum cromossomo da espécie, recebendo a denominação de *Uncharacterized chromosome* (Unchr) (SCHOCH *et al.*, 2020). A ferramenta do PCR eletrônico é um procedimento computacional utilizado para se visualizar o tamanho do produto de uma PCR e localizar a região do genoma que está sendo analisada (ROTMISTROVSKY *et al.*, 2004). Em virtude desses *primers* serem inicialmente

desenvolvidos para a espécie *C.arabica*, essa ferramenta permite verificar se o par de *primers* amplifica mais de uma região do genoma.

Além de utilizar marcadores distribuídos em diferentes cromossomos, na seleção dos *primers* deste estudo, também foram consideradas as regiões do genoma que são amplificadas. Foram utilizados *primers* que se encontram em regiões codantes (CaEST-002, CaEST-029, CaEST-071 e CaEST-103) e em regiões não codantes (SSR 07, SSR 16, SSR 34, SSR 119 e SSR 122). A diferença entre esses marcadores está na forma que foram desenvolvidos. Os *primers* CaEST são oriundos de sequências de DNA complementares às sequências de RNA mensageiro, representando parte de uma região transcrita do genoma em certa condição fisiológica (SOUSA *et al.*, 2017). Dessa forma, esses marcadores amplificam regiões codantes do genoma. Os demais marcadores foram desenvolvidos em regiões aleatórias do genoma. A escolha de *primers* que se encontram em diferentes cromossomos, aliados a *primers* de regiões codantes e não codantes garantiu a eficiência dos mesmos na discriminação e caracterização dos cafeeiros em estudo.

A média de bandas obtidas por *primer* (5,16) indica uma grande variabilidade genética entre os genótipos da população de estudo. Isso pode ser explicado pelo modo de reprodução da espécie (polinização cruzada), formação de populações com alta variabilidade genética e elevados níveis de heterozigose apresentados pela espécie *C. canephora* (FERRÃO *et al.*, 2013).

2.4.2 Diversidade genética e perfil molecular

Uma vez obtido o perfil molecular único de cada genótipo do BAG e do programa de melhoramento da Embrapa Rondônia, observou-se que todos os cafeeiros foram discriminados, validando a unicidade de cada acesso (Tabela S1). Essa é uma informação relevante principalmente ao se considerar os clones selecionados pelos próprios agricultores e que apesar de serem amplamente cultivados, ainda são desconhecidos em muitos aspectos. Essas informações configuram-se como ferramenta auxiliar aos descritores recomendados para a espécie *C. canephora*, sendo de grande utilidade em testes de distinguibilidade, homogeneidade e estabilidade, que são exigidos no processo de registro e proteção de cultivares, bem como na manutenção da diversidade no programa de melhoramento (ALKIMIM *et al.*, 2020) e rastreabilidade do material genético.

A caracterização molecular realizada auxiliou, ainda, o conhecimento do *background* genético dos clones, quanto à similaridade com as variedades botânicas de *C. canephora*. Os cafeeiros dessa espécie, cultivados no Brasil, podem ser divididos em duas variedades botânicas de acordo com o seu centro de diversidade e suas características adaptativas. Apesar das diferenças fenotípicas existentes entre as duas variedades botânicas Conilon e Robusta, a classificação dos cafeeiros não é simples, visto que a forma de reprodução alógama da espécie propicia plantas com elevada amplitude fenotípica e heterozigose (FERRÃO *et al.*, 2019). Além disso, deve se considerar a ocorrência de cruzamentos naturais entre as variedades botânicas que podem dificultar essa discriminação. Nesse contexto, o uso de marcadores moleculares pode auxiliar na discriminação dos cafeeiros, uma vez que permitem diferenciar polimorfismo a nível de DNA e são neutros fenotipicamente (ALKIMIM *et al.*, 2021).

No presente trabalho, foram analisados clones sabidamente da variedade Conilon e da variedade Robusta, os quais foram classificados quando introduzidos em Rondônia. Os cafeeiros Conilon, quando comparados com Robusta, apresentaram menor diversidade entre os indivíduos dentro de variedade botânica e maior diversidade entre variedade botânica. Pelo dendrograma da Figura 1, observou-se a formação de três grupos de Conilon e cinco grupos de Robusta. Dos 12 clones de Conilon analisados, a maioria (C9-BAG, C6-BAG, C8-BAG, C2-BAG, C7-BAG e C5-BAG) ficou alocado em um único grupo (Con 1), confirmando essa menor diversidade entre esses genótipos. Segundo Charrier & Berthaud, 1985 a menor diversidade da variedade botânica Conilon também é observada no centro de origem de cafeeiro, que tem sua origem em centro de diversidade entre o Gabão, próximo ao rio Kouilou, onde ficou restrito (CHARRIER & BERTHAUD, 1985). Além disso, sua introdução no Brasil foi realizada com poucos genótipos no Estado do Espírito Santo, onde seu cultivo ainda é predominante (FERRÃO *et al.*, 2007). SOUZA *et al.* (2013) e FERRÃO *et al.* (2013), em trabalhos anteriores com populações de *C. canephora*, também encontraram resultados semelhantes, indicando a menor variabilidade apresentada pelos cafeeiros Conilon.

Os genótipos Conilon C1-BAG, C3-BAG e C12- BAG ficaram alocados nos grupos de Robusta, indicando possíveis hibridações naturais. No contexto de uso de marcadores moleculares em auxílio a programas de melhoramento, além de permitir a discriminação das plantas pelo seu *fingerprint* e de auxiliar o melhorista na hibridação e seleção de plantas, os marcadores também permitem identificar genótipos que apesar de suas características predominantes em campo, fazem parte de diferentes populações de melhoramento..

Para melhor classificar os genótipos em suas populações de melhoramento, foi utilizada classificação probabilística a partir do coeficiente de ancestralidade (Figura 2). Os resultados dessa análise também indicam que os cafeeiros C-1 BAG e C-3 BAG são provavelmente híbridos naturais provenientes do cruzamento entre as duas variedades botânicas, uma vez que apresentam em seu genoma grande porcentagem do genoma Robusta.

Já os cafeeiros Robusta apresentaram uma maior diversidade formando cinco grupos (ROB 1, ROB2, ROB 3, ROB4 e ROB 5) com pouca similaridade entre os genótipos. O centro de diversidade da variedade botânica Robusta corresponde a uma grande região na África central e é estruturado em cinco subgrupos, SG1, SG2, B, C e um grupo de acessos selvagens chamado de UW (NG *et al.*, 2017) o que pode estar associado a maior variabilidade observada no presente trabalho, fora do centro de origem desse cafeeiro. Além disso, os cafeeiros Robusta não ficaram restritos ao seu local de origem, sendo amplamente disseminados para outros países, (ANAGBOGU *et al.*, 2019). Essa maior variabilidade pode ser observada nas duas metodologias utilizadas no presente estudo: o dendrograma (Figura 1), cujos cafeeiros apresentam pouca similaridade entre si e a formação de cinco grupos distintos, a classificação baseada no coeficiente de ancestralidade (Figura 2), onde a maior parte dos genótipos Robusta não mostrou probabilidade de pertencimento aos mesmos grupos da variedade botânica Conilon.

A caracterização molecular de maior número de acessos do programa de melhoramento (Figura 3) corroborara a hipótese em que a identidade de alguns cafeeiros Conilon foram questionadas. O acesso C1-BAG apresentou alta similaridade com a cultivar BRS 2357, que corresponde a um híbrido entre Conilon e Robusta. O cafeeiro C1-BAG continuou alocado em um subgrupo de Robustas.

Nessa mesma caracterização, híbridos naturais e direcionados foram analisados e, como esperado, esses cafeeiros apresentaram alta variabilidade e foram alocados nos diferentes grupos formados no dendrograma. A exceção foi no grupo VIII, onde ficou formado apenas por Robusta (R13-BAG e R15-BAG). Esses dois acessos se diferenciaram de todos os genótipos híbridos, indicando que possuem pouca similaridade com os híbridos e as cultivares registradas. Essa variabilidade pode ser explorada no melhoramento genético, incluindo esses robustas como genitores. Oliveira *et al.*, 2018 obtiveram as maiores estimativas de ganho de seleção a partir da hibridação entre genitores divergentes das variedades botânicas Conilon e Robusta. (MONTAGNON *et al.*, 1998) Ainda na década de 80 (MONTAGNON *et al.*, 1998) também observaram maiores potenciais de ganhos a partir

da hibridação entre plantas diferentes. Híbridos naturais de menor porte e resistência ao déficit hídrico do Conilon, com um maior peneira e resistência a pragas e doenças do Robusta tem se destacado nas avaliações de campo. A formação de diversos subgrupos, contendo cafeeiros Robusta, Conilon e híbridos, (Figuras 3, 4 e 5) indicam a variação genética entre acessos foi maior que a variação observada dentro das variedades botânicas. Esse resultado foi confirmado na análise molecular de variância (AMOVA), onde 80% da variância se encontra no polimorfismo entre os acessos e não entre os grupos (Tabela 5). A maior variabilidade entre acessos pode ser explicada pelo fato das duas variedades botânicas, Conilon e Robusta, apresentarem características divergentes e complementares, favorecendo a seleção de indivíduos híbridos (ALKIMIM *et al.*, 2020). Segundo ROCHA *et al.*, 2021 a seleção praticada pelos agricultores favoreceu a seleção dos indivíduos híbridos com características complementares das variedades botânicas associadas a expressão do vigor do híbrido, que hoje, caracterizam a cafeicultura na Amazônia Ocidental

Os dados obtidos no presente trabalho mostram, também, que os cafeeiros plantados, atualmente, no Estado de Rondônia apresentam maior introgressão do genoma Robusta (Figura 6), o que pode ser resultante de maior adaptação dessa variedade botânica na região. ESPINDULA *et al.* (2018) também relataram que essa variedade botânica é predominante e mais disseminada no Estado. No entanto, estudos realizados com cafeeiros introduzidos na década de 90 mostraram que esses apresentavam maior introgressão de Conilon (FERRÃO *et al.*, 2013; SOUZA *et al.*, 2013). Esses estudos anteriores, realizados com populações de *C. canephora* e com os mesmos *primers* utilizados no presente estudo, sugeriram também que os cafeeiros inicialmente plantados em Rondônia apresentavam menor variabilidade e agrupamentos mais homogêneos entre os grupos varietais. Segundo SOUZA *et al.* (2013), a menor variabilidade entre os acessos explorados na Amazônia e no melhoramento genético, no início do cultivo de *C. canephora* na região poderia limitar a sustentabilidade da cultura, de forma que os pesquisadores sugeriram que novos genótipos fossem introduzidos no programa de melhoramento do Estado. Com a distribuição de sementes da variedade botânica Robusta realizada pela Embrapa na década de 80 juntamente com o resultado do trabalho dos próprios cafeicultores que participaram ativamente desse processo de seleção, , atualmente, observa-se um parque cafeeiro mais diverso. Os dados do presente estudo evidenciam a maior variabilidade atual dos cafeeiros, cuja a condução e cruzamento podem aumentar o rendimento das lavouras da região. Além disso, as análises moleculares e estatísticas

realizadas permitiram indicar quais os cafeeiros podem ser hibridizados, visando o máximo da diversidade e, conseqüentemente, potencial de maiores ganhos genéticos.

Além de avaliar acessos do BAG e do programa de melhoramento genético dos cafeeiros da Amazônia, foram analisados clones de domínio público que são cultivados pelos produtores. Esses clones pertencem a lavouras de cafeeiros selecionados pelos próprios produtores. A história da cafeicultura no Estado se iniciou com a introdução de ambas as variedades botânicas. No entanto, devido ao modo de reprodução da espécie, híbridos naturais foram formados e passaram a ser preferidos e selecionados pelos produtores da região, principalmente por apresentarem características intermediárias entre as duas variedades (PEREIRA *et al.*, 2000). Os produtores selecionaram e clonaram os cafeeiros mais adaptados e que apresentavam fenótipos superiores para formarem suas lavouras, sendo iniciado no Estado a adoção de lavouras de origem clonal fundamentada no cultivo de clones altamente produtivos e tolerantes às doenças do cafeeiro (PEREIRA *et al.*, 2000). Assim, os clones de domínio público amplamente cultivados no Estado não apresentam passaportes genéticos, dificultando o entendimento da variabilidade desses materiais e uso mais eficientes dos cafeeiros. Desse modo, o presente estudo priorizou entender e classificar, por meio de análises moleculares e estatísticas, os principais clones plantados pelos produtores. Os clones dos produtores foram comparados com acessos sabidamente da variedade Conilon e da variedade Robusta, bem como híbridos obtidos por cruzamentos controlados entre as duas variedades. Os dados mostraram uma maior similaridade desses cafeeiros com os acessos híbridos (F_{st} 0,019) (Tabela 4), que corrobora com o processo de origem das lavouras dos produtores, os quais foram selecionados através de características intermediárias entre as duas variedades botânicas. Na figura 5 é possível observar clones com alta similaridade aos híbridos como os acessos H12-BAG e o clone N8, BRS 2314 e o clone AS12, H11-BAG e o clone 25 e o H15-BAG e o clone L1. Em campo, essas plantas também apresentam características fenotípicas intermediárias, típicas de híbridos entre Robusta e Conilon, como porte médio, estabilidade de produção, frutos de tamanho médio, ausência de incidência de ferrugem e robustez. Também podem ser observados clones com alta similaridade com a variedade Conilon, como os clones LB010, clone R 22, clone N12e clone 21 e com a variedade Robusta (clone 41, clone 06, LB015, clone AS7). A genotipagem dos clones de produtores e as análises estatísticas realizadas no presente estudo permitiram discriminar e caracterizar os 38 cafeeiros e contabilizar a porcentagem de introgressão do genoma de cada cafeeiro com base nos grupos varietais Conilon e Robusta (Figura 6). Como descrito anteriormente, pelo fato de a

cafeicultura rondoniense apresentar, atualmente, características e genótipos predominantes da variedade Robusta, é possível observar que os maiores valores de similaridade são entre os clones dos produtores e a variedade Robusta. No entanto, destacam-se os clones L1 e AS5 que ficaram alocados separadamente no grupo III da Figura 4.

2.4 Conclusão

Por meio das análises realizadas, foi selecionado um conjunto de marcadores moleculares capazes de discriminar com eficiência cafeeiros da espécie *C. canephora*.

Foram construídos perfis moleculares para os 96 clones de *C. canephora*, úteis para rastreabilidade dos acessos e cultivares.

Os marcadores moleculares e os grupos de diversidade identificados e validados nesse trabalho podem ser utilizados nos programas de melhoramento, a fim de auxiliar e acelerar etapas de seleção e cruzamento.

Referências

- Akpertey A, Padi FK et al (2021) Effectiveness of Single Nucleotide Polymorphism markers in genotyping germplasm collections of *Coffea canephora* using KASP assay. **Frontiers in plant science**, 11: 612593. <https://doi.org/10.3389/fpls.2020.612593>
- Alkimim ER, Caixeta ET, Sousa TV et al (2020) Selective efficiency of genome-wide selection in *Coffea canephora* breeding. **Tree Genetics & Genomes**, 16(3):1-11. <https://doi.org/10.1007/s11295-020-01433-3>
- Alves EA, de Souza CA, Rocha RB et al (2020) Efeito da fermentação sobre qualidade da bebida do Café Robusta (*Coffea canephora*) cultivado na Amazônia Ocidental. **Revista Ifes Ciência** 6:159-170. <https://doi.org/10.36524/ric.v6i3.875>
- Anagbogu CF, Bhattacharjee R, Ilori C et al (2019) Genetic diversity and re-classification of coffee (*Coffea canephora* Pierre ex A. Froehner) from South Western Nigeria through genotyping-by-sequencing-single nucleotide polymorphism analysis. **Genetic Resources and Crop Evolution**, 66(3):685-696. <https://doi.org/10.1007/s10722-019-00744-2>
- Araújo ES, Arriel NHC, dos Santos RC et al (2019) Assessment of genetic variability in sesame accessions using SSR markers and morpho-agronomic traits. **Australian Journal of Crop Science** 13: 45-54. <https://doi.org/10.21475/ajcs.19.13.01.p1157>
- Benti T, Gebre E, Tesfaye K et al (2021) Genetic diversity among commercial arabica coffee (*Coffea arabica* L.) varieties in Ethiopia using simple sequence repeat markers. **Journal of Crop Improvement** 35:147-168. <https://doi.org/10.1080/15427528.2020.1803169>
- Cenci A, Combes MC, Lashermes, P (2012) Genome evolution in diploid and tetraploid *Coffea* species as revealed by comparative analysis of orthologous genome segments. **Plant molecular biology**, 78(1): 135-145. <https://doi.org/10.1007/s11103-011-9852-3>
- Charrier A, Berthaud, J (1985) Botanical classification of coffee. In *Coffee*. Springer, Boston, MA, pp 13-47
- Combes MC, Andrzejewski S, Anthony F et al (2000) Characterization of microsatellite loci in *Coffea arabica* and related coffee species. **Molecular ecology** 9:1178-1180. <https://doi.org/10.1046/j.1365-294X.2000.00954-5.x>
- COMPANHIA NACIONAL DE ABASTECIMENTO (2022) **Acompanhamento da Safra Brasileira: Café Boletim da Safra 2021 de café**. Acessado em 20 de junho 2022
- Da Fonseca AFA, Sedyama T, Cruz CD et al (2006) Divergência genética em café conilon. **Pesquisa Agropecuária Brasileira** 41:599-605. <https://doi.org/10.1590/S0100-204X2006000400008>
- Dangl GS, Yang J, Golino DA et al (2009) A practical method for almond cultivar identification and parental analysis using simple sequence repeat markers. **Euphytica** 168:

41–48. <https://doi.org/10.1007/s10681-008-9877-0>

Davis AP (2011) *Psilanthus mannii*, the type species of *Psilanthus*, transferred to *Coffea*. **Nordic Journal of Botany** 29(4): 471-472. <https://doi.org/10.1111/j.1756-1051.2011.01113.x>

Diniz LE, Sakiyama NS, Lashermes P et al (2005) Analysis of AFLP markers associated to the Mex-1 resistance locus in Icatu progenies. **Crop Breed Appl Technol** 5(4): 387. <https://doi.org/10.12702/1984-7033.v05n04a031>

Dubberstein D, Partelli FL, Guilhen JHS et al (2020) Biometric traits as a tool for the identification and breeding of *Coffea canephora* genotypes. **Genet Mol Res**. <https://doi.org/10.4238/gmr18541>

Espindula MC, Araújo LFBD, Schmidt R et al (2020) Early induction of orthotropic shoots in *Coffea canephora*. **Revista Ceres** 67: 281-287. <https://doi.org/10.1590/0034-737X202067040005>

Ferrão LFV, Caixeta ET, Souza FDF et al (2013) Comparative study of different molecular markers for classifying and establishing genetic relationships in *Coffea canephora*. **Plant systematics and evolution**, 299(1):225-238. Disponível em: <https://doi.org/10.1007/s00606-012-0717-2>

Ferrão LFV, Ferrão RG, Ferrão MAG et al (2019) Accurate genomic prediction of *Coffea canephora* in multiple environments using whole-genome statistical models. **Heredity** 122(3): 261-275. <https://doi.org/10.1038/s41437-018-0105-y>

Ferrão MAG, Ferrão RG, Fonseca AD et al (2007) Origem, dispersão geográfica, taxonomia e diversidade genética de *Coffea canephora*. **Café conilon**. Incaper, Vitória, pp 66–91

Ferrão RG (2004) *Biometria aplicada ao melhoramento genético do café Conilon*. Tese, Universidade Federal de Viçosa.

Ferrão LFV, Caixeta ET, Souza FDF et al (2013) Comparative study of different molecular markers for classifying and establishing genetic relationships in *Coffea canephora*. **Plant systematics and evolution** 299(1): 225-238. <https://doi.org/10.1007/s00606-012-0717-2>

GAVA, Romário *et al.* **Café Conilon**. [S. l.: s. n.], 2017.

Goudet J (2005) Hierfstat, a package for R to compute and test hierarchical F-statistics. **Molecular ecology notes** 5(1):184-186. <https://doi.org/10.1111/j.1471-8286.2004.00828.x>

Hassan DA, Hama-Ali EO (2022) Evaluation of gene flow and genetic diversity in rice accessions across Kurdistan region-iraq using SSR markers. **Molecular biology reports**, 49(2):1007-1016. <https://doi.org/10.1007/s11033-021-06920-x>

Huded AKC, Jingade P, Bychappa M et al (2020) Genetic diversity and population structure analysis of *Coffee (Coffea canephora)* germplasm collections in Indian Gene Bank employing SRAP and SCoT markers. **International Journal of Fruit Science**, 20:S757-

S784. <https://doi.org/10.1080/15538362.2020.1768618>

Krishna Kumar S, Feldman MW et al (2016) Limitations of GCTA as a solution to the missing heritability problem. **Proceedings of the National Academy of Sciences** 113(1):E61-E70. <https://doi.org/10.1073/pnas.1520109113>

Lashermes P, Andrzejewski S, Bertrand B et al (2000) Molecular analysis of introgressive breeding in *Coffee* (*Coffea arabica* L.). **Theoretical and applied genetics** 100(1): 139-146. <https://doi.org/10.1007/s001220050019>

Lashermes P, Combes MC, Prakash NS et al (2001) Genetic linkage map of *Coffea canephora*: effect of segregation distortion and analysis of recombination rate in male and female meioses. **Genome** 44(4):589-595. <https://doi.org/10.1139/gen-44-4-589>

Louthan AM, Kay KM (2011) Comparing the adaptive landscape across trait types: larger QTL effect size in traits under biotic selection. **BMC evolutionary biology**, 11(1): 1-12. <https://doi.org/10.1186/1471-2148-11-60>

Merot-l'Anthoene V, Tournebize R, Darracq O et al (2019) Development and evaluation of a genome-wide *Coffee* 8.5 K SNP array and its application for high-density genetic mapping and for investigating the origin of *Coffea arabica* L. **Plant biotechnology journal**, 17:1418-1430. <https://doi.org/10.1111/pbi.13066>

Montes LD (2009) Molecules That Changed the World (KC Nicolaou and T. Montagnon). **Journal of Chemical Education** [s. l.] <https://doi.org/10.1021/ed086p1372>

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.

Nei M, Li WH (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. **Proceedings of the National Academy of Sciences** 76:5269-5273. <https://doi.org/10.1073/pnas.76.10.5269>

Paradis E (2010) pegas: an R package for population genetics with an integrated–modular approach. **Bioinformatics** 26:419-420. <https://doi.org/10.1093/bioinformatics/btp696>

Pereira RDCA, Léo, FJDS et al (2000) Avaliação de genótipos dos cafeeiros Arabica e Robusta na Amazônia Ocidental. **Acta Amazonica** 30:535-535. <https://doi.org/10.1590/1809-43922000304541>

Ramalho AR, Rocha RB, Souza FF et al (2016) Progresso genético da produtividade de café beneficiado com a seleção de clones de cafeeiro 'Conilon'. **Revista Ciência Agrônômica** 47:516-523. <https://doi.org/10.5935/1806-6690.20160062>

Rambaut A (2016) FigTree–Tree Figure Drawing Tool, version 1.4. 3. Molecular evolution, phylogenetics and epidemiology.

Resende MDV (2016) Software Selegen-REML/BLUP: a useful tool for plant breeding. **Crop Breed Appl Biotechnol** 16:330–339. <https://doi.org/10.1590/1984-70332016v1 6n4a49>

Rotmistrovsky K, Jang W et al (2004) A web server for performing electronic PCR. **Nucleic acids research**, 32:W108-W112. <https://doi.org/10.1093/nar/gkh450>

Semagn K, Bjørnstad Å, Ndjioudjop MN (2006) An overview of molecular marker methods for plants. **African journal of biotechnology**, 5(25). <https://doi.org/10.5897/AJB2006.000-5110>

Schoch CL, Ciufu S, Domrachev M et al (2020) NCBI Taxonomy: a comprehensive update on curation, resources and tools. Database [*S. l.: s. n.*]. <https://doi.org/10.1093/database/baaa062>

Sousa TV, Caixeta ET, Alkimim ER et al (2017) Molecular markers useful to discriminate *Coffea arabica* cultivars with high genetic similarity. **Euphytica** 213(3):1-15. <https://doi.org/10.1007/s10681-017-1865-9>

Souza FDF, Caixeta ET, Ferrão LFV et al (2013) Molecular diversity in *Coffea canephora* germplasm conserved and cultivated in Brazil. **Crop breeding and applied biotechnology**, 13:221-227. <https://doi.org/10.1590/s1984-70332013000400001>

Zhao H, Li Q, Li J et al (2006) The study of neighboring nucleotide composition and transition/transversion bias. **Science in China Series C: Life Sciences**, 49(4):395-402.

3 Capítulo II

ESTRUTURA POPULACIONAL E DIVERSIDADE GENÉTICA DO BANCO DE GERMOPLASMA DE *Coffea canephora* AVALIADO COM MARCADORES SSR E SNP

Resumo

Dados sobre a diversidade genética e estrutura populacional de bancos de germoplasma são de fundamental importância para o melhoramento genético das plantas. Nesse estudo objetivou-se comparar a eficiência de uso dos marcadores SSR e SNP para esses estudos de diversidade genética em cafeeiros da espécie *Coffea canephora*. Foram analisados 165 acessos do banco de germoplasma do 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). A população analisada foi constituída por 42 genótipos da variedade botânica Conilon, 41 genótipos da variedade botânica Robusta, e 82 híbridos interpopulacionais obtidos por meio de cruzamentos artificiais. Esses acessos foram genotipados com 20 marcadores SSR e 17.885 marcadores SNP. O número médio de alelos por loco e a diversidade gênica (D) foram maiores para os marcadores SSR do que para os SNP. As estimativas da distância de Roger e probabilidade de pertencimento de cada acesso ao seu respectivo grupo baseados em metodologias bayesianas mostrou uma maior eficiência dos marcadores SNP para caracterização de acessos Conilon e dos marcadores SSR para Robusta. As análises de componentes principais mostraram semelhanças entre ambos os marcadores formando três grupos homogêneos e alguns acessos que não puderam ser classificados. Esses resultados sugerem que as mesmas conclusões podem ser extraídas de ambos os marcadores. No entanto, um menor número de marcadores SSR permitiu classificar a população em estudo, apresentando eficiência na análise de diversidade desses acessos de *C. canephora*.

Palavras-chave: Melhoramento do cafeeiro. Distância modificada de Rogers. Índice F_{ST} . Metodologias bayesianas.

Abstract

Data on the genetic diversity and population structure of germplasm banks are of fundamental importance for the genetic improvement of plants. The objective of this study was to compare the efficiency of the use of SSR and SNP indicators for the genetic diversity and population structure of the *C. canephora* germplasm bank. A total of 165 accessions from the breeding program Empresa de Pesquisa Agropecuária de Minas Gerais (Epamig) in partnership with the Federal University of Viçosa UFV) and the Empresa Brasileira de Pesquisa Agropecuária - Café (Embrapa Café) were analyzed. The population is composed of clones of the botanical varieties Conilon and Robusta and of interpopulation hybrids (Conilon x Robusta). The botanical Conilon had 42 genotypes, the Robusta variety consisted of 41 genotypes, in addition to 82 interpopulation hybrids. These accessions were genotyped with 20 SSR markers and 17,885 SNP markers. The mean number of alleles per locus and genetic diversity (D) were higher for SSR markers than for SNP. Estimates of Roger's distance and the probability of each accession belonging to its respective group in Bayesian methodologies were greater using SNP markers for the Conilon variety and SSR markers for the Robusta variety. Regardless of marker type, the principal component analysis formed three main groups though some accessions were not assigned to any groups. These results suggest that the same results can be observed for both can marker types. However, a smaller number SSR markers provides great informativeness for these diversity analysis.

Keywords: Coffee breeding. Modified Rogers distance. F_{ST} index. Bayesian methodologies.

3.1 Introdução

Mais de 7 milhões de plantas cultivadas são mantidas em acessos de banco de germoplasma em todo mundo. No entanto, estratégias para quantificar, entender e utilizar essa diversidade podem ser fatores limitantes que resultam na falta de uma visão clara e integrada dos acessos que essas coleções contêm (MILNER *et al.*, 2019). A maior parte das informações atuais desses bancos de germoplasma incluem vários formatos de dados, dados coletados manualmente e curados de formas inerentemente difíceis de serem analisadas e comparadas dentro e entre as coleções (LANGRIDGE & WAUGH, 2019). Nesse contexto, a identificação precisa das plantas em qualquer coleção de germoplasma é importante para facilitar o seu manejo e uso nos programas de melhoramento genético. Assim, a utilização de marcadores moleculares de baixo custo e confiáveis são ferramentas eficientes para auxiliar o gerenciamento de um banco de germoplasma (AKPERTEY *et al.*, 2021).

O banco de germoplasma de Minas Gerais mantido em parceria pela Empresa de Pesquisa Agropecuária de Minas Gerais (Epamig), Universidade Federal de Viçosa (UFV) e Empresa Brasileira de Pesquisa Agropecuária (Embrapa) dispõe de uma coleção de genótipos da espécie *Coffea canephora*. Esse banco é constituído por exemplares das variedades botânicas Conilon e Robusta, bem como híbridos provenientes do cruzamento entre essas variedades. Devido ao fato das sementes de café serem recalcitrantes, não sendo adequadas para o armazenamento (AKPERTEY *et al.*, 2021), os acessos são representados por plantas propagadas por estacas (acessos clonais), mantidas em campo. Esse modo de conservação pode resultar em identificações errôneas, perda de identificação, redundâncias e outros erros no passaporte dos acessos do banco de germoplasma. Nesse sentido, como as características morfológicas são geralmente influenciadas por fatores ambientais, os marcadores moleculares são ferramentas poderosas para estimar a diversidade genética presente nesses acessos (DE JESÚS PIRES *et al.*, 2022). Diferentes marcadores moleculares foram desenvolvidos para serem usados para análise de cafeeiros, em especial os marcadores microssatélites ou SSR (*Simple Sequence Repeat*) (COMBES *et al.*, 2000; BARUAH *et al.*, 2003; PONCET *et al.*, 2004.; AGGARWAL *et al.*, 2007; HENDRE *et al.*, 2008; MISSIO *et al.*, 2009 ; AGGARWAL *et al.*, 2007; FERRÃO *et al.*, 2015) A utilidade desses marcadores tem sido descrita em diferentes pesquisas em *C. canephora*, incluindo diversidade genética (KIWUKA *et al.*, 2021 ; DE SOUZA *et al.*, 2021 ; DE AQUINO *et al.*, 2022) caracterização de germoplasma (BHAT *et al.*, 2005; SOUZA *et al.*, 2013; NG *et al.*, 2017), mapas de ligação

(COULIBALY *et al.*, 2003; CUBRY *et al.*, 2008; FERRÃO *et al.*, 2013). Embora a utilidade desses marcadores seja de grande auxílio no estudo genético e no manejo do germoplasma de *C. canephora*, esses marcadores ainda não foram automatizados para a genotipagem em larga escala, afetando o seu rendimento (ZHOU *et al.*, 2016).

Os marcadores de polimorfismo de nucleotídeo único (SNP – *Single Nucleotide Polymorphism*), por sua vez, são marcadores adequados para análise em larga escala genômica (MEROT-L'ANTHOENE *et al.*, 2019). Em comparação com os marcadores SSR, os marcadores SNP são mais abundantes e de alto rendimento (MEROT-L'ANTHOENE *et al.*, 2019), no entanto, são menos informativos quando comparados aos SSR. Dadas as vantagens e desvantagens de cada marcador, suas utilidades em diferentes análises devem ser comparadas e consideradas. O tipo de marcador molecular a ser usado deve ser adequado ao objetivo da pesquisa, uma vez que cada tipo de marcador é baseado em diferentes princípios, diferindo quanto a consistência dos dados, tempo de obtenção de resultados, nível de polimorfismo, custo e facilidade de implementação e sua reprodutibilidade (AKPERTEY *et al.*, 2021).

Assim, o presente estudo objetivou comparar a eficiência de uso dos marcadores SSR e SNP para a diversidade genética e estrutura populacional de acessos do banco de germoplasma de *C. canephora*.

3.2 Material e métodos

3.2.1 Material vegetal

Um total de 165 genótipos de *C. canephora*, pertencente ao Banco de Germoplasma (BAG) de Minas Gerais, foi analisado com marcadores moleculares. Esse BAG tem sido mantido em uma parceria entre a Empresa de Pesquisa Agropecuária de Minas Gerais (Epamig), Universidade Federal de Viçosa (UFV) e Empresa Brasileira de Pesquisa Agropecuária - Café (Embrapa Café). Esses materiais genéticos são mantidos na área experimental situada no distrito São José do Triunfo, pertencente a cidade de Viçosa, MG. O plantio dos cafeeiros foi realizado em espaçamento de 3,0 m entre as fileiras e 1,5 m entre as plantas, disponibilizados em delineamento em blocos casualizados (DBC).

Os acessos analisados correspondem a clones dos grupos varietais Conilon e Robusta e por híbridos interpopulacionais (Conilon x Robusta). A variedade botânica Conilon (Tabela 1) foi constituído por 42 genótipos, obtidos do Instituto Capixaba de Pesquisa, Assistência

Técnica e Extensão Rural (Incaper). A variedade botânica Robusta (Tabela 2) foi constituído por 41 génotipos obtidos do Centro Agronómico Tropical de Investigación y Enseñanza (CATIE). Além desses, 82 híbridos interpopulacionais foram obtidos por meio de cruzamentos artificiais avaliados a partir de um dialelo parcial formado por cinco genitores masculinos Conilon e cinco genitores femininos Robusta (Tabela 3).

Tabela 1. Acessos de cafeeiro pertencentes a variedade botânica Conilon.

Amostra	Nº Registro das Plantas	Amostra	Nº Registro das Plantas
1	UFV 513	25	UFV 3628-40
2	UFV 3627-20	26	UFV 3628-42
3	UFV 3627-24	27	UFV 3628-44
4	UFV 3627-27	28	UFV 3628-45
5	UFV 3627-29	29	UFV 3628-46
6	UFV 3627-30	30	UFV 3629-4
7	UFV 3627-31	31	UFV 3629-10
8	UFV 3628-1	32	UFV 3629-11
9	UFV 3628-2	33	UFV 3629-17
10	UFV 3628-3	34	UFV 3629-24
11	UFV 3628-4	35	UFV 3629-25
12	UFV 3628-5	36	UFV 3629-26
13	UFV 3628-10	37	UFV 3629-27
14	UFV 3628-16	38	UFV 3629-28
15	UFV 3628-17	39	UFV 3629-29
16	UFV 3628-22	40	UFV 3629-30
17	UFV 3628-23	41	UFV 3629-31
18	UFV 3628-24	42	UFV 3629-34
19	UFV 3628-26		
20	UFV 3628-27		
21	UFV 3628-28		
22	UFV 3628-29		
23	UFV 3628-32		
24	UFV 3628-35		

Tabela 2. Acessos de cafeeiro pertencentes a variedade botânica Robusta.

Amostra	Nº Registro das Plantas	Amostra	Nº Registro das Plantas
1	UFV 514	29	UFV 3376-8
2	UFV 3356-71	30	UFV 3376-9
3	UFV 3357-93	31	UFV 3377-12
4	UFV 3358-88	32	UFV 3630-2
5	UFV 3360-169	33	UFV 3630-05
6	UFV 3360-171	34	UFV 3630-06
7	UFV 3361-148	35	UFV 3630-07
8	UFV 3361-151	36	UFV 3630-10
9	UFV 3362-118	37	UFV 3631-09
10	UFV 3363-125	38	UFV 3631-10
11	UFV 3365-144	39	UFV 3631-11
12	UFV 3366-134	40	UFV 3631-13
13	UFV 3366-138	41	UFV 3356-76
14	UFV 3366-139		
15	UFV 3367-96		
16	UFV 3367-97		
17	UFV 3367-101		
18	UFV 3367-105		
19	UFV 3368-52		
20	UFV 3368-58		
21	UFV 3370-50		
22	UFV 3371-19		
23	UFV 3371-20		
24	UFV 3373-36		
25	UFV 3373-43		
26	UFV 3374-29		
27	UFV 3375-65		
28	UFV 3375-66		

Tabela 3. Acessos de cafeeiro pertencentes aos Híbridos resultantes do cruzamento entre as variedades botânicas Robusta x Conilon.

Amostra	Nº Registro das Plantas	Amostra	Nº Registro das Plantas
1	H0921-1	47	H094-5
2	H0916-1	48	H0915-5
3	H0920-1	49	H0914-5
4	H093-1	50	H0912-6
5	H0915-1	51	H0911-6
6	H098-2 A	52	H099-6
7	H0913-1	53	H0914-6
8	H0918-1	54	H0910-6
9	H0912-1	55	H0913-7
10	H0917-1	56	H0911-7
11	H099-1	57	H0910-7
12	H098-1	58	H0912-7
14	H0922-1	59	H099-7
15	H0919-2	60	H099-8
16	H0914-2	61	H092-8
17	H0911-2	62	H0912-8
18	H0916-2	63	H097-8
19	H099-2	64	H0910-8
20	H0918-2	65	H095-9
21	H098-2 B	66	H0910-9
22	H094-2	67	H092-10
23	H092-2	68	H099-10
24	H0913-2	69	H093-10
25	H0915-2	70	H095-11
26	H0917-2	71	H093-11
27	H098-3	72	H094-12
28	H0912-3	73	H093-12
29	H0917-3	74	H097-13
30	H0911-3	75	H098-13
31	H0918-3	76	H094-13
32	H0913-3	77	H095-14
33	H0914-3	79	H097-16
34	H0915-3	80	H097-17
35	H0914-4	81	H097-18
36	H0911-4	82	H094-18
37	H0910-4		
38	H0913-4		
39	H0912-4		
40	H0915-4		
42	H0916-4		
43	H0913-5		
44	H098-5		
45	H0910-5		
46	H0911-5		

3.2.2 Extração do DNA genômico

Seguindo a metodologia proposta por DINIZ *et al.* (2005), folhas jovens dos 165 cafeeiros tiveram seu DNA genômico extraídos no Laboratório de Biotecnologia do Cafeeiro – Biocafé/UFV. Após a extração, a qualidade do DNA foi verificada em gel de agarose 1% e sua concentração verificada em *NanoDrop 2000*. As concentrações das amostras foram padronizadas e posteriormente armazenadas.

3.2.3 Genotipagem com marcadores SSR

Foram utilizados 20 pares de *primers* SSR (Tabela 4).

Tabela 4 – Primer SSR utilizados na genotipagem dos 165 acessos de *C. canephora* pertencentes ao banco de germoplasma da UFV

<i>Primer</i>	<i>Forward primer (5' → 3')</i>	<i>Reverse primer (5' → 3')</i>
SSR-07 ¹	TGACATAGGGGGCTAAATTG	TTAATGGTGACGCTTTGATG
SSR-08 ¹	CACTGGCATTAGAAAGCACC	GGCAAAGTCAATGATGACTC
SSR-13 ¹	TGGCCGTGATAATAAACAGC	ATGTGGCAATCTAAAGCCAA
SSR-16 ¹	ACCCGAAAGAAAGAACCAAG	CCACACAACCTCTCCTCATTC
SSR-21 ¹	GACCATTACATTTACACACAC	GCATTTTGTTCACACTGTA
SSR-29 ²	GGCTTCTTGGGTGTCTGTGT	CCATTGGCTTTGTATTTCTGG
SR-30 ³	ATGGGGCCAACCTTGAATATG	CAGGGCATCTATCTACTTCTCTTT
SSR-34 ³	GGAGACGCAGGTGGTAGAAG	TCGAGAAGTCTTGGGGTGTT
SSR-35 ³	CTGGCATTAGAAAGCACCTTG	GCTTGGCTCACTGTAGGACTG
SSR-37 ³	CAACACTATCTCTTGATTTTTCACT	CGTGCAAGTCACATACTTTACTAC
SSR-39 ²	TCCCCCATCTTTTTCTTTCC	GGGAGTGTTTTTGTGTTGCTT
SSR-40 ³	AAAGGAAAATTGTTGGCTCTGA	TCCACATACATTTCCCAGCA
SSR-42 ²	TTGCTTGCTTGTCTGTTAT	TGACACGAGAGTTAGAAATGA
SSR-43 ³	TTTTCTGGGTTTTCTGTGTTCTC	TAACTCTCCATTCCC GCATT
SSR-46 ³	AATGAAGAAGAGGGTGGTG	CGAGGGTATTGTTTTCCAG
SSR-48 ²	AGCAAGTGGAGCAGAAGAAG	CGGTGAATAAGTCGCAGTC
SSR-49 ³	TGGAGAAGGCTGTTGAAACC	GGCGTGAAGCAAAAAGGTAT
SSR-52 ³	GACCAAATGTCAGCTCATTG	GCCGACTGCTCTTTTAGTGT
SSR-55 ²	GCAGGTATTTGAAGGATGAACC	GTGTAGGTGGTGCGATGTGT
SSR-56 ³	AGCTATCTTTATCTCACACACACA	GTTAGTGTTCGATTTGGTACTG

Referências: ¹ Combes *et al.* (2000); ² Coulibaly *et al.* (2003); ³ Poncet *et al.* (2004)

A amplificação por PCR foi realizada da seguinte forma: 50ng de DNA, 0,6 unidades de Taq DNA polimerase (Invitrogen), tampão enzimático 1x, 1mM de MgCl₂, 150μM de cada dNTP e 0,1μM de cada *primer*, completando o volume total de 20 μl com água milli-Q estéril. As reações foram realizadas nos termocicladores PTC200 (MJ Research) e Veriti (Applied Biosystems). Após a desnaturação inicial a 94°C por 2 minutos, foram realizados 13 ciclos de PCR constituídos por desnaturação a 94°C por 30 segundos, anelamento com temperatura decrescente de 1°C em cada ciclo (67 a 55°C) por 30 segundos e extensão a 72°C por 30 segundos. Em seguida, foram realizados mais 30 ciclos compostos por desnaturação a 94°C, anelamento a 55°C e extensão a 72°C por 30 segundos em cada estágio. A extensão final foi realizada a 72°C por 8 minutos. Os produtos resultantes da reação de PCR foram separados por eletroforese em gel desnaturante em poliacrilamida a 6% e visualizados por coloração com nitrato de prata (BRITO *et al.*, 2010). Os fragmentos amplificados por marcadores SSR (padrão codominante) foram codificados de acordo com cada genótipo e seguindo os requisitos para as análises estatísticas.

3.2.4 Genotipagem com marcadores SNP

A genotipagem por SNP foi realizada por meio da metodologia *CaptureSeq* da empresa *RAPiD Genomics*, localizada na Florida/EUA. Nessa metodologia, inicialmente foram construídas sondas a partir de sequências de referência provenientes do Projeto Brasileiro do Genoma Café e do banco de dados do genoma da espécie *C. canephora*. De posse dessas sondas, os genótipos foram sequenciados via plataforma *Illumina* e os marcadores identificados por abordagem de bioinformática. Os marcadores SNP obtidos foram analisados quanto à qualidade, eliminando-se os alelos menos frequente (MAF) e os marcadores que apresentaram grandes quantidades de dados perdidos (*CR-call rate*). Os valores usados para eliminação foram MAF (*Minor allele frequency*) inferior a 0,01 e CR inferiores a 90%. A codificação dos dados para a confecção da matriz de marcadores SNP foram (0,1 e 2), em que 0 indica que os indivíduos eram homozigotos para o alelo menos frequente, 1 indica indivíduos heterozigotos e 2 indica que os indivíduos eram homozigotos para o alelo mais frequente.

3.2.5 *Análises estatísticas*

Todas as análises descritas abaixo foram realizadas para os marcadores SSR e SNP. O número de alelos por locos, o número de alelos por grupo, a heterozigozidade observada (H_o) e esperada (H_e), e a diversidade gênica D (NEI, Masatoshi, 1972) foram determinados para cada variedade botânica e para todos os indivíduos. Os valores estatísticas F (Coeficiente de fixação de Wright's (F_{st}) e Coeficiente de parentesco (F_{is})) foram obtidos de acordo com Wright (1965). As médias e os intervalos da distância de Roger modificada (DRM) (WRIGHT, 1946) foram calculadas dentro e entre os grupos varietais. As estatísticas F foram calculadas por meio do pacote “hierfstat” (GOUDET, 2005) e as demais com o pacote “adegenet” (JOMBART, 2008). Para a análise da estrutura populacional dos acessos de cafeeiro utilizou-se a análise discriminante de componentes principais (DAPC), um método bayesiano projetado para identificar e descrever agrupamentos de indivíduos geneticamente relacionados. Essa metodologia permite extrair informações avançadas de dados genéticos, fornecendo atribuições a grupos e contribuição de alelos individuais para a estruturação da população. Essa metodologia utiliza o *software* STRUCTURE como *benchmark* (JOMBART *et al.*, 2010). A variância amostral da Distância modificada de Roger (DMR) foi estimada a partir de um subconjunto de marcadores (5, 10, 15, 20, 40, 60, 80 e 100% do conjunto total de marcadores), aleatoriamente selecionados. Com base nos marcadores selecionados, a DMR e o coeficiente de variação (CV) foram calculados. Todos os cálculos foram realizados usando o *software* R v 4.1.3 (R Development Core Team 2022).

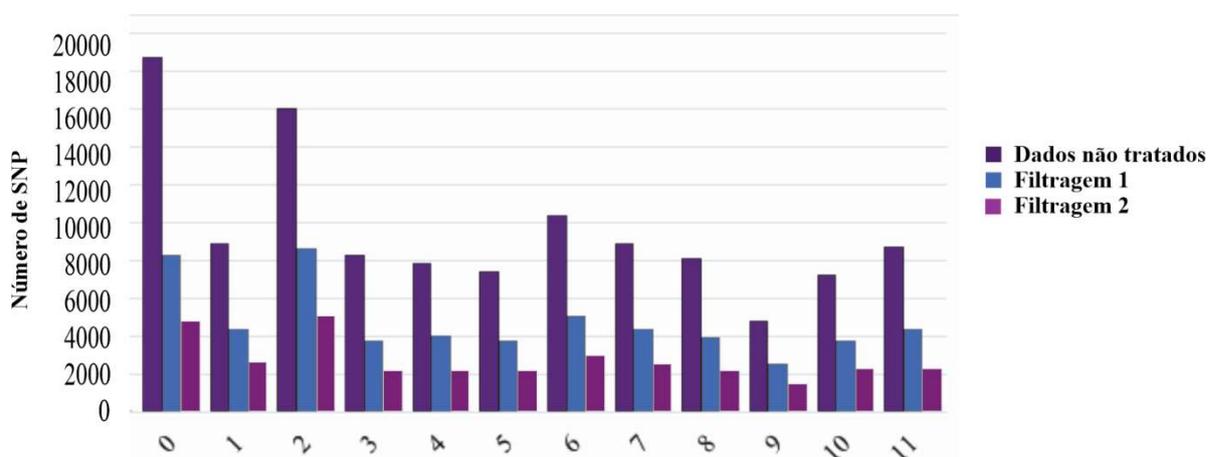
3.3 Resultados

3.3.1 *Comparação entre estimativas de diversidade genética para os marcadores SSR e SNP*

Os 165 cafeeiros da BAG de Minas Gerais foram analisados com 20 marcadores SSR e 17.885 marcadores SNP. Foram obtidos, inicialmente, 18.111 SNP, no entanto, foram realizadas duas filtragens de qualidade sendo a $MAF \geq 0,01$ e $CR \geq 90\%$ utilizada para realizar as demais análises (Figura 1).

A distribuição física dos SNP no genoma de *C. canephora* foi relativamente uniforme, apresentando poucas lacunas e estando uniformemente distribuídos em todos os 11 cromossomos da espécie (Figura 1). Foram observados 23 SNP alocados em um conjunto de *Scaffolds* não ordenados que não foram agrupados em nenhum cromossomo da espécie recebendo a denominação de *Uncharacterized chromosome* (Unchr) (SCHÖCH *et al.*, 2020). Esses SNP foram descartados das análises (Figura 1).

O número médio de alelos por loco, considerando todos os 165 cafeeiros da BAG, foi



de 4,4 para os marcadores SSR e 2,0 para os SNP (Tabela 5). Os marcadores também foram analisados considerando os cafeeiros divididos em grupos segundo a variedade botânica que pertencem.

Figura 1. Marcadores SNP distribuídos ao longo dos 11 cromossomos de *C. canephora*. O eixo x representa os 11 cromossomos das espécies e o eixo y representa o número de SNP. As barras em diferentes tonalidades indicam a quantidade de SNP distribuídos pelo genoma antes e após as filtragens de qualidade. Filtragem 1: $MAF \geq 0,05$ e $CR \geq 95\%$; filtragem 2: $MAF \geq 0,01$ e $CR \geq 90\%$.

A diversidade D total, considerando todos os cafeeiros, foi de 0,70 para os SSR e de 0,27 para os SNP. As estimativas de D por variedade botânica variaram de 0,67 (SSR) e 0,28 (SNP) para Conilon, 0,69 (SSR) e 0,27 (SNP) para Robusta e 0,59 (SSR) e 0,26 (SNP) para os Híbridos. O índice geral de fixação F_{ST} foi de 0,105 e 0,002 para os marcadores SSR e SNP, respectivamente (Tabela 5).

Os dados moleculares (SSR e SNP) foram analisados considerando três grupos de acordo com as variedades botânicas nos quais os acessos pertencem. Os marcadores SSR para DMR entre os grupos variou de 0,02 (Conilon e Híbridos) a 0,03 (Robusta e Híbridos). Em comparação, para os SNP, a distância média variou de 0,08 (Conilon e Híbridos) a 0,09 (Robusta e Híbrido) (Tabela 6).

Tabela 5. Média e variação do número de alelos por locos, diversidade gênica (D) e F_{ST} para 165 genótipos de *C. canephora* pertencentes as variedades botânicas Conilon, Robusta e os Híbridos resultantes desse cruzamento. Classificação baseada nos fenótipos apresentados pelos acessos.

	Grupos Varietais			Total
	Conilon (42)	Robusta (41)	Híbrido (82)	
SSR				
N. de alelo por grupo	20	19	22	-
Média alelo por loco	4	3,8	4,4	4,4
Ho	0,59	0,67	0,52	0,59
He	0,67	0,69	0,59	0,65
Divers. (D)	0,67	0,69	0,59	0,7
F_{st}	0,105	0,205	0,075	0,128
F_{is}	0,11	0,02	0,1	0,08
SNP				
Média alelo por loco	1,99	1,99	1,98	2
Ho	0,2612	0,2164	0,2365	0,2381
He	0,2836	0,2742	0,2666	0,2748
Divers. (D)	0,2833	0,2734	0,2663	0,2731
F_{st}	-0,029	0,005	0,031	0,002
F_{is}	0,07	0,21	0,11	0,13

SSR: Marcadores microssatélites, SNP: Marcadores de polimorfismo de nucleotídeo único; Ho: heterogeneidade observada, He: Heterogeneidade esperada; D: Diversidade gênica; F_{st} : Correlação dos genes dos indivíduos dentro de uma população; F_{is} : Correlação dos genes entre indivíduos de uma população em relação ao conjunto das populações.

Tabela 6. Média das estimativas de distância de Roger modificadas aos pares calculadas a partir de marcadores SSR (acima da diagonal) e marcadores SNP (abaixo da diagonal) entre os grupos de cafeeiros alocados segundo a variedade botânica o qual pertence (grupos Conilon, Robusta e Híbridos).

	Conilon	Robusta	Híbrido
Conilon	—	0,0269	0,0299
Robusta	0,0895	—	0,0339
Híbrido	0,0831	0,0997	—

3.3.2 Descrição da estrutura populacional

Para um melhor entendimento da estrutura populacional dos acessos de germoplasma, todas as análises foram realizadas em grupos baseados nos fenótipos apresentados pelos acessos, sendo eles Conilon, Robusta e Híbridos (provenientes dos cruzamentos entre as duas variedades). Na Tabela 7 estão apresentadas as porcentagens de pertencimento baseada na análise de estrutura bayesiana. Para os marcadores SNP, a variedade Conilon apresentou 80,9% de pertencimento, alocando 34 acessos corretamente como Conilon, 6 acessos como Robusta e 2 acessos como Híbridos. Para a variedade Robusta, foi observada uma taxa de pertencimento de 48,8% com 20 acessos alocados corretamente como Robusta, 8 acessos como Conilon e 13 acessos como Híbridos. Enquanto a variedade híbrida apresentou 92,7 % de pertencimento, alocando 76 acessos corretamente como Híbridos, 5 acessos como Robusta e 1 acesso como Conilon.

Ao se observar a mesma análise realizada para os marcadores SSR, a variedade Conilon apresentou 73,8 % com 31 acessos alocados corretamente e 6 acessos alocados no grupo Robusta e 5 Híbridos respectivamente. A variedade Robusta apresentou 68,3% de pertencimento com 28 acessos alocados corretamente e 6 alocados nos grupos Conilon e 7 Híbridos respectivamente. Já a variedade Híbrida apresentou 85,4% de pertencimento com 70 acessos alocados corretamente, 8 acessos no grupo Robusta e 4 acessos no grupo Conilon. A média da porcentagem de acerto salientou a maior eficiência dos marcadores SSR (75,8%) em relação aos marcadores SNP (74,1%).

Tabela 7. Atribuição dos genótipos das variedades botânicas identificados com base na probabilidade de associação mais alta para marcadores de repetição de sequência simples (SSR) e polimorfismo de nucleotídeo único (SNP) para os grupos varietais Conilon, Robusta e Híbridos.

SNP	Conilon	Robusta	Híbrido	Total	PPG
Conilon	34	6	2	42	80,9%
Robusta	8	20	13	41	48,8%
Híbrido	1	5	76	82	92,7%
Média					74,1%
SSR					
Conilon	31	6	5	42	73,8%
Robusta	6	28	7	41	68,3%
Híbrido	4	8	70	82	85,4%
Média					75,8%

PPG: porcentagem de pertencimento a grupo.

No PCA (Análise de Componentes Principais) (Figura 2) com base nas estimativas da distância de Rogers de todos os 165 acessos de *C. canephora*, a primeira e a segunda coordenada principal (CP) explicaram 67 e 7,84 % da variância molecular para o marcador SSR e 63,9 e 7,69% para o marcador SNP. Para ambos os marcadores, CP1 e CP2, separaram claramente os três clusters, demonstrando semelhanças entre as informações das variedades botânicas

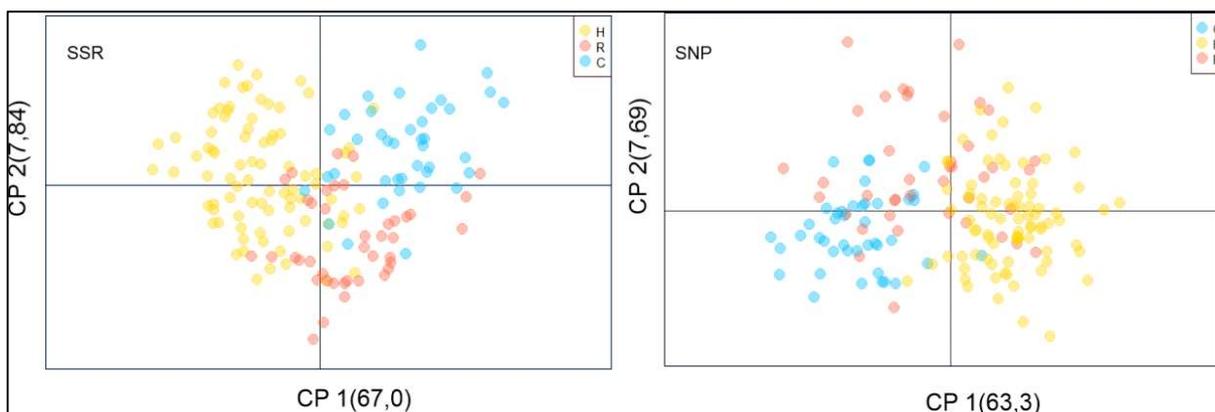


Figura 2. Análise de coordenadas principais de 165 acessos com base na distância de Roger modificada calculada a partir de 20 SSR (esquerda) ou de 17.885 SNP (direita) loci marcadores. Os genótipos foram atribuídos ao subgrupo de acordo com a probabilidade máxima de associação. CP1 e CP2 são a primeira e a segunda coordenada principal, respectivamente, e o número entre parênteses se refere à proporção de variância explicada pelas coordenadas principais

O coeficiente de variação (CV) (Figura 3) das estimativas da distância de Rogers aumentou exponencialmente com a diminuição do número de marcadores SSR e SNP. Para os marcadores SSR (azul) observou-se que ao se atingir 50% do total de marcadores (10) o CV atinge um equilíbrio e começa a diminuir exponencialmente, indicando que 10 marcadores SSR já seriam suficientes para a diferenciação desses acessos. Enquanto para os marcadores SNP o platô é atingido com 8942 marcadores.

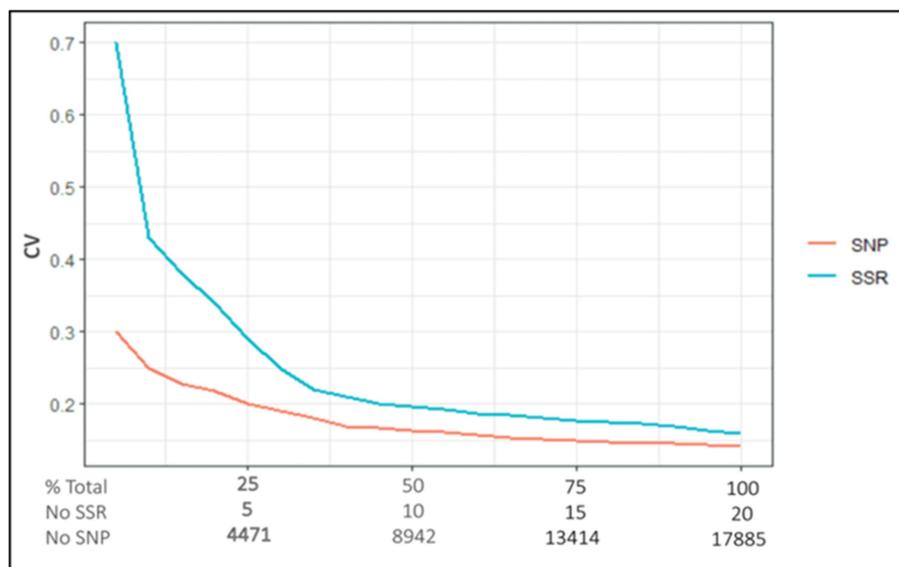


Figura 3. O coeficiente médio de variação (CV) da distância de Roger modificada (MRD) estimado através de 165 acessos de *C. canephora*. Em vermelho estão representados os marcadores SNP e em azul os marcadores SSR.

3.4 Discussão

3.4.1 Comparação entre estimativas de diversidade genética para os marcadores SSR e SNP

O número médio de alelos por loco foi maior para os marcadores SSR (4,4) do que para os SNP (2,0) (Tabela 5). Esse resultado está de acordo com o tipo de marcador usado, visto que os SSR são multialélicos (SOUZA *et al.*, 2013) e os SNP são, em geral, bialélicos (VIGNAL *et al.*, 2002). Valores mais altos para a diversidade gênica (D) também foram observados quando os cafeeiros foram analisados com marcadores SSR (0,70), enquanto os marcadores SNP apresentaram 0,27. A máxima diversidade gênica (D) que marcadores bialélicos, como é o caso dos SNP, podem apresentar é 0,50, enquanto marcadores

multialélicos, como os SSR, é de 1,0 (VAN INGHELANDT *et al.*, 2010), explicando a diferença obtida nos dados do presente estudo. Além disso, os maiores valores encontrados para os SSR podem ser resultantes da seleção de marcadores polimórficos. Dos mais de 400 marcadores SSR desenvolvidos para café (AGGARWAL *et al.*, 2007; BARUAH *et al.*, 2003; FERRÃO, *et al.*, 2015; LASHERMES *et al.*, 2000; MISSIO *et al.*, 2010; PONCET *et al.*, 2004), foram utilizados um conjunto (20 marcadores) sabidamente mais polimórficos. Os marcadores SSR aqui utilizados vieram de prévias seleções realizadas por trabalhos de caracterização molecular e diversidade genética em populações de *C. canephora*. (MISSIO *et al.*, 2010; FERRÃO *et al.*, 2013; SOUZA *et al.*, 2013). Para os SNP, foram utilizados todos os marcadores obtidos no ensaio *CaptureSeq*, sem selecionar os mais polimórficos.

Quando comparados os valores de D (Tabela 5) calculados para SSR e SNP dentro dos grupos Conilon, Robusta e Híbrido, observou-se tendências diferentes entre os marcadores. Para os marcadores SSR os menores valores de D são observados nos Híbridos seguida das variedades Conilon e Robusta. Enquanto para os marcadores SNP os valores entre as variedades botânicas Conilon (0,28) é maior que para a variedade Robusta (0,27). Esses resultados indicam que os dois tipos de marcadores diferem quanto a estimar a diversidade genética do germoplasma em estudo.

O índice geral de fixação F_{ST} (também chamado de migração gênica ou fluxo de alelos) é uma medida útil na diferenciação populacional, pois estima a correlação dos alelos entre indivíduos de uma população em relação ao conjunto de populações (KHAN *et al.*, 2021). A interpretação desse índice foi realizada de acordo com WRIGHT (1946), onde os valores de F_{ST} podem variar de 0 a 1. O valor 0 significa compartilhamento completo de material genético e 1 significa que toda variação genética é explicada pela estrutura populacional. Uma escala padrão de índice de fixação de 0 a 0,05, como o caso dos SNP (0,02) (Tabela 5), indica uma pequena diferenciação e variação de 0,15 a 0,25, como o caso dos SSR (0,12) (Tabela 5), indica diferenciação genética moderada. No caso da análise população considerando acessos separados em grupos de acordo com o fenótipo, percebe-se por meio do índice F_{ST} , os marcadores SSR foram mais efetivos em diferenciar essa população.

Outra análise realizada foi a estimativa de distância modificada de Rogers (ROGERS, 1972) (Tabela 6). Essa estimativa tem sido mais recomendada para marcadores codominantes, como os SSR e SNP, além de apresentar a vantagem de ser uma modificação da distância Euclidiana, que corresponde a uma média em relação a todos os locos em estudo. Essa

estratégia facilita a comparação dos valores obtidos sem penalizar os marcadores bialelicos como o caso dos SNP (BOCCACCI *et al.*, 2021). Com o uso dessa metodologia, verificou-se que, existe diferença nas estimativas da distância de Rogers quando comparados as mesmas variedades botânicas, como por exemplo Conilon e Robusta, com marcadores SSR (0,02) e SNP (0,08). Essa correlação imperfeita entre as estimativas de distância se deve provavelmente ao fato de que a taxa de mutação apresentada pelos marcadores moleculares SSR é consideravelmente maior que a dos marcadores SNP (VAN INGHELANDT *et al.*, 2010)). Os SSR geralmente são resultado da adição ou deleção de unidades repetidas, portanto o polimorfismo observado pelos indivíduos são diferenças nesses números de repetição, que são causados pelo deslizamento da polimerase na replicação do DNA ou por erros de recombinação. Assim, mutações que escapam da correção pelo sistema de incompatibilidade do DNA formam novos alelos nos loci SSR, o que significa que esse marcadores são mais informativos, mas também mais sujeitos a erros de mutação que os marcadores SNP (VIEIRA *et al.*, 2016).

3.4.2 Descrição da estrutura populacional

Manutenção e conservação de bancos de germoplasma são essenciais para o desenvolvimento do melhoramento genético de plantas. Para utilizar eficientemente os acessos do banco no melhoramento e explorá-lo ao máximo é preciso conhecê-los detalhadamente e obter informações sobre a estrutura populacional desse germoplasma, bem como a diversidade genética (AKPERTEY *et al.*, 2021).

Para as análises de diversidade do banco de germoplasma do presente trabalho, os acessos foram divididos em três grupos segundo seus fenótipos (Conilon, Robusta e Híbridos). Por meio da abordagem baseada na análise bayesiana (Tabela 7), observou-se que, independente dos parâmetros utilizados para a atribuição dos acessos a suas variedades botânicas, alguns acessos não foram atribuídos aos seus respectivos grupos. A Tabela 7 exemplifica também qual marcador foi mais eficiente em classificar cada acesso em sua respectiva variedade botânica, ao se comparar as médias das probabilidades de pertencimento de grupo (PPG), os marcadores SSR tiveram um melhor resultado, indicando que segundo essa metodologia os marcadores SSR se mostraram mais eficientes em classificar a estrutura populacional desses acessos.

A variedade botânica Conilon, em virtude de sua dispersão e centro de diversidade serem restritos ao Gabão (CHARRIER & BERTHAUD, 1985) e sua introdução no Brasil ser realizada com poucos genótipos (FERRÃO *et al.*, 2007), apresenta uma menor diversidade quando comparada a variedade Robusta. Estudos realizados por SOUZA *et al.* (2013) com marcadores SSR também descreveram a menor variabilidade apresentada pelos Conilon, assim esses genótipos por serem mais similares entre si foram melhores agrupados com os marcadores SNP (Tabela 7). Entretanto, quando se trata da variedade Robusta que apresenta uma maior variabilidade em seus acessos, os marcadores SSR, por serem multialélicos, conseguiram captar e classificar melhor essa variabilidade sendo os mais adequados para essa variedade. Quando analisamos as variedades híbridas, por se tratar de acessos oriundos dessas duas variedades botânicas com características complementares, ambos os marcadores conseguiram detectar e classificar a variabilidade presente nesses acessos.

Os resultados da análise de componentes principais (Figura 2) válida os resultados encontrados nas análises de estrutura de população, onde é possível observar que a melhor classificação dos acessos Conilon é pelos marcadores SNP e a melhor classificação dos acessos Robusta pelos marcadores SSR. É possível observar ainda acessos que não conseguiram ser atribuídos aos seus respectivos grupos, esse resultado pode ser explicado pelo modo de reprodução da espécie *C. canephora*, indicando que esses genótipos possam se tratar de híbridos naturais entre as duas variedades, ou mesmo podem se tratar de acessos com a identificação errônea.

Além das estimativas de distância genética e diversidade populacional, no presente estudo também foi utilizada a variância associada a aplicação desses marcadores, de modo a concluir qual o marcador mais eficaz na população em análise. Para isso foi comparado o CV com as estimativas de distância de Roger. Foi observado um CV maior para os marcadores SSR do que para os marcadores SNP (Figura 3). Pode-se verificar também que ambos os marcadores atingem um platô, indicando que acima de um certo número de marcadores a precisão obtida com esses marcadores adicionais estava diminuindo. Para os marcadores SSR, esse platô é de 10 marcadores, enquanto para os marcadores SNP são 8.942. Estudos realizados por Garcia *et al.* (2004) corroboram esses resultados, indicando que a genotipagem com mais do que um certo número de marcadores melhora apenas marginalmente a precisão das estimativas de diversidade da população. Ambos os marcadores atingiram a equiparidade com 50% do total de seus marcadores, no entanto os marcadores SSR mostraram maiores valores de CV atingido com poucos marcadores, fato que pode minimizar custos e

ferramentas computacionais. A análise de marcadores SNP por se tratarem de dados de alta densidade exigem ferramentas computacionais robustas, além de amplo conhecimento estatístico para a interpretação desses dados (RAFALSKI, 2002), os resultados obtidos demonstram que um conjunto de 8.942 SNP já seriam suficientes para a o estudo da estrutura genética dessa população de *C. canephora*.

Quando analisamos em conjunto ambos os marcadores, é preciso ponderar além do CV das análises os custos de genotipagem envolvidos. Com base nos dados de genotipagem, 10 marcadores SSR seriam necessários para atingir uma precisão adequada para esse estudo de diversidade enquanto seriam necessários 8.942 marcadores SNP. Além desses dados do CV, as posições no genoma desses marcadores também influenciam. Neste estudo, todos os 11 cromossomos da espécie possuem um marcador molecular SSR, assim a precisão dos marcadores SSR se tornou maior por esses marcadores serem escolhidos e não distribuídos aleatoriamente no genoma. Dessa forma, com base nos resultados obtidos, sugere-se que sejam utilizados 800 vezes mais marcadores SNP do que SSR para o estudo desse germoplasma.

Em síntese, ambos os marcadores moleculares foram eficientes para quantificar a diversidade e a estrutura genética dos acessos do banco de germoplasma de *C. canephora* do Estado de Minas Gerais. No entanto, entre as duas variedades botânicas foi possível observar que a variedade Conilon, que apresenta uma base genética mais estreita, foi mais bem classificada com os marcadores bialélicos (SNP) enquanto os acessos Robusta, que apresentam uma base genética maior, foram melhores classificados com os marcadores multialélicos (SSR). Já os acessos Híbridos apresentaram pouca divergência nos resultados entre os dois marcadores.

3.5 Conclusões

Os resultados desse estudo indicaram que as mesmas conclusões sobre a estrutura e a diversidade genética dos acessos do banco de germoplasma de *C. canephora* podem ser obtidas de ambos os tipos de marcadores, SSR e SNP. No entanto, um menor número de marcadores SSR são eficientes em classificar a população em estudo. Ressalta-se que, no presente estudo, os marcadores SSR foram previamente selecionados, sendo utilizado apenas um conjunto sabidamente polimórfico, o que não foi realizado para os marcadores SNP.

3.6 Referências

Akpertey A, Padi FK et al (2021) Effectiveness of Single Nucleotide Polymorphism markers in genotyping germplasm collections of *Coffea canephora* using KASP assay. **Frontiers in plant science**, 11: 612593. <https://doi.org/10.3389/fpls.2020.612593>

Alkimim ER, Caixeta ET, Sousa TV et al (2020) Selective efficiency of genome-wide selection in *Coffea canephora* breeding. **Tree Genetics & Genomes**, 16(3):1-11. <https://doi.org/10.1007/s11295-020-01433-3>

Alves EA, de Souza CA, Rocha RB et al (2020) Efeito da fermentação sobre qualidade da bebida do Café Robusta (*Coffea canephora*) cultivado na Amazônia Ocidental. **Revista Ifes Ciência** 6:159-170. <https://doi.org/10.36524/ric.v6i3.875>

Anagbogu CF, Bhattacharjee R, Ilori C et al (2019) Genetic diversity and re-classification of coffee (*Coffea canephora* Pierre ex A. Froehner) from South Western Nigeria through genotyping-by-sequencing-single nucleotide polymorphism analysis. **Genetic Resources and Crop Evolution**, 66(3):685-696. <https://doi.org/10.1007/s10722-019-00744-2>

Araújo ES, Arriel NHC, dos Santos RC et al (2019) Assessment of genetic variability in sesame accessions using SSR markers and morpho-agronomic traits. **Australian Journal of Crop Science** 13: 45-54. <https://doi.org/10.21475/ajcs.19.13.01.p1157>

Benti T, Gebre E, Tesfaye K et al (2021) Genetic diversity among commercial arabica coffee (*Coffea arabica* L.) varieties in Ethiopia using simple sequence repeat markers. **Journal of Crop Improvement** 35:147-168. <https://doi.org/10.1080/15427528.2020.1803169>

Cenci A, Combes MC, Lashermes, P (2012) Genome evolution in diploid and tetraploid *Coffea* species as revealed by comparative analysis of orthologous genome segments. **Plant molecular biology**, 78(1): 135-145. <https://doi.org/10.1007/s11103-011-9852-3>

Combes MC, Andrzejewski S, Anthony F et al (2000) Characterization of microsatellite loci in *Coffea arabica* and related coffee species. **Molecular ecology** 9:1178-1180. <https://doi.org/10.1046/j.1365-294X.2000.00954-5.x>

Dangl GS, Yang J, Golino DA et al (2009) A practical method for almond cultivar identification and parental analysis using simple sequence repeat markers. **Euphytica** 168: 41–48. <https://doi.org/10.1007/s10681-008-9877-0>

Davis AP (2011) *Psilanthus mannii*, the type species of *Psilanthus*, transferred to *Coffea*. **Nordic Journal of Botany** 29(4): 471-472. <https://doi.org/10.1111/j.1756-1051.2011.01113.x>

Diniz LE, Sakiyama NS, Lashermes P et al (2005) Analysis of AFLP markers associated to the Mex-1 resistance locus in Icatu progenies. **Crop Breed Appl Technol** 5(4): 387. <https://doi.org/10.12702/1984-7033.v05n04a031>

Dubberstein D, Partelli FL, Guilhen JHS et al (2020) Biometric traits as a tool for the identification and breeding of *Coffea canephora* genotypes. **Genet Mol Res**. <https://doi.org/10.4238/gmr18541>

Ferrão LFV, Caixeta ET, Souza FDF et al (2013) Comparative study of different molecular markers for classifying and establishing genetic relationships in *Coffea canephora*. *Plant systematics and evolution*, 299(1):225-238. Disponível em: <https://doi.org/10.1007/s00606-012-0717-2>

Ferrão LFV, Ferrão RG, Ferrão MAG et al (2019) Accurate genomic prediction of *Coffea canephora* in multiple environments using whole-genome statistical models. **Heredity** 122(3): 261-275. <https://doi.org/10.1038/s41437-018-0105-y>

Ferrão MAG, Ferrão RG, Fonseca AD et al (2007) Origem, dispersão geográfica, taxonomia e diversidade genética de *Coffea canephora*. **Café conilon. Incaper**, Vitória, pp 66–91

Ferrão RG (2004) Biometria aplicada ao melhoramento genético do café Conilon. Tese, Universidade Federal de Viçosa.

Ferrão LFV, Caixeta ET, Souza FDF et al (2013) Comparative study of different molecular

markers for classifying and establishing genetic relationships in *Coffea canephora*. **Plant systematics and evolution** 299(1): 225-238. <https://doi.org/10.1007/s00606-012-0717-2>

Goudet J (2005) Hierfstat, a package for R to compute and test hierarchical F-statistics. **Molecular ecology notes** 5(1):184-186. <https://doi.org/10.1111/j.1471-8286.2004.00828.x>

Hassan DA, Hama-Ali EO (2022) Evaluation of gene flow and genetic diversity in rice accessions across Kurdistan region-iraq using SSR markers. **Molecular biology reports**, 49(2):1007-1016. <https://doi.org/10.1007/s11033-021-06920-x>

Huded AKC, Jingade P, Bychappa M et al (2020) Genetic diversity and population structure analysis of *Coffee (Coffea canephora)* germplasm collections in Indian Gene Bank employing SRAP and SCoT markers. **International Journal of Fruit Science**, 20:S757-S784. <https://doi.org/10.1080/15538362.2020.1768618>

Krishna Kumar S, Feldman MW et al (2016) Limitations of GCTA as a solution to the missing heritability problem. **Proceedings of the National Academy of Sciences** 113(1):E61-E70. <https://doi.org/10.1073/pnas.1520109113>

Lashermes P, Andrzejewski S, Bertrand B et al (2000) Molecular analysis of introgressive breeding in *Coffee (Coffea arabica L.)*. **Theoretical and applied genetics** 100(1): 139-146. <https://doi.org/10.1007/s001220050019>

Lashermes P, Combes MC, Prakash NS et al (2001) Genetic linkage map of *Coffea canephora*: effect of segregation distortion and analysis of recombination rate in male and female meioses. **Genome** 44(4):589-595. <https://doi.org/10.1139/gen-44-4-589>

Louthan AM, Kay KM (2011) Comparing the adaptive landscape across trait types: larger QTL effect size in traits under biotic selection. **BMC evolutionary biology**, 11(1): 1-12. <https://doi.org/10.1186/1471-2148-11-60>

Merot-l'Anthoene V, Tournebize R, Darracq O et al (2019) Development and evaluation of a

genome-wide *Coffea* 8.5 K SNP array and its application for high-density genetic mapping and for investigating the origin of *Coffea arabica* L. **Plant biotechnology journal**, 17:1418-1430. <https://doi.org/10.1111/pbi.13066>

Montes LD (2009) Molecules That Changed the World (KC Nicolaou and T. Montagnon). **Journal of Chemical Education** [s. l.] <https://doi.org/10.1021/ed086p1372>

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.

Paradis E (2010) pegas: an R package for population genetics with an integrated–modular approach. **Bioinformatics** 26:419-420. <https://doi.org/10.1093/bioinformatics/btp696>

Pereira RDCA, Léo, FJDS et al (2000) Avaliação de genótipos dos cafeeiros Arabica e Robusta na Amazônia Ocidental. **Acta Amazonica** 30:535-535. <https://doi.org/10.1590/1809-43922000304541>

Ramalho AR, Rocha RB, Souza FF et al (2016) Progresso genético da produtividade de café beneficiado com a seleção de clones de cafeeiro 'Conilon'. **Revista Ciência Agronômica** 47:516-523. <https://doi.org/10.5935/1806-6690.20160062>

Rambaut A (2016) FigTree–Tree Figure Drawing Tool, version 1.4. 3. Molecular evolution, phylogenetics and epidemiology.

Resende MDV (2016) Software Selegen-REML/BLUP: a useful tool for plant breeding. **Crop Breed Appl Biotechnol** 16:330–339. <https://doi.org/10.1590/1984-70332016v1 6n4a49>

Rotmistrovsky K, Jang W et al (2004) A web server for performing electronic PCR. **Nucleic acids research**, 32:W108-W112. <https://doi.org/10.1093/nar/gkh450>

Semagn K, Bjørnstad Å, Ndjiondjop MN (2006) An overview of molecular marker methods for plants. **African journal of biotechnology**, 5(25). <https://doi.org/10.5897/AJB2006.000-5110>

Schoch CL, Ciufu S, Domrachev M et al (2020) NCBI Taxonomy: a comprehensive update on curation, resources and tools. Database [S. l.: s. n.]. <https://doi.org/10.1093/database/baaa062>

Sousa TV, Caixeta ET, Alkimim ER et al (2017) Molecular markers useful to discriminate *Coffea arabica* cultivars with high genetic similarity. **Euphytica** 213(3):1-15. <https://doi.org/10.1007/s10681-017-1865-9>

Souza FDF, Caixeta ET, Ferrão LFV et al (2013) Molecular diversity in *Coffea canephora* germplasm conserved and cultivated in Brazil. **Crop breeding and applied biotechnology**, 13:221-227. <https://doi.org/10.1590/s1984-70332013000400001>

620

621 **Tabelas Suplementares**

Tabela S1. Perfil molecular dos 96 cafeeiros em estudo. A codificação de cada indivíduo se encontra na Tabela 1. Os 12 *primers* microsatélites utilizados nas análises se encontram em cada coluna. (continua)

Pop	SSR07	SSR16	SSR35	SSR34	C02	C29	C71	C103	SAT235	SAT207	SSR119	SSR122
1	11	22	11	22	44	33	22	35	22	13	13	33
2	54	33	45	34	44	56	-9	-9	22	44	13	-9
3	45	23	24	23	22	55	13	-9	11	14	23	45
4	11	33	45	55	22	54	23	35	11	36	13	23
5	33	33	25	46	35	44	22	34	11	34	13	33
6	11	22	55	66	44	56	-9	11	22	33	33	24
7	64	33	24	26	44	54	23	35	11	34	23	24
8	45	33	45	22	44	44	11	55	11	14	13	33
9	11	22	45	26	-9	-9	22	11	11	-9	13	45
10	54	22	52	36	26	66	22	33	22	34	13	-9
11	11	33	45	22	22	35	23	35	11	34	11	23
12	11	22	42	26	22	55	13	33	11	46	11	33
13	11	33	35	45	36	55	23	35	-9	34	13	23
14	11	33	25	26	44	44	13	35	11	46	23	22
15	54	22	25	23	44	56	23	33	11	23	11	34
16	45	33	45	22	-9	55	22	44	11	36	23	24
17	23	33	45	25	22	55	22	34	11	33	23	45
18	55	22	44	55	33	44	22	33	11	33	11	35
19	35	22	25	25	22	55	22	35	11	34	23	33

622

623

624

625

Tabela S1. Perfil molecular dos 96 cafeiros em estudo. A codificação de cada indivíduo se encontra na Tabela 1. Os 12 *primers* microsatélites utilizados nas análises se encontram em cada coluna. (continuação)

20	33	33	55	25	22	66	33	45	11	33	13	33
21	54	33	45	45	44	56	22	43	22	34	11	33
22	54	22	44	26	44	33	23	34	11	13	13	23
23	54	34	45	23	22	55	13	44	11	36	13	35
24	11	22	54	22	24	45	33	44	11	13	13	45
25	66	22	25	25	66	55	23	35	11	34	13	23
26	45	32	45	34	44	66	12	44	22	24	11	55
27	33	23	24	25	22	64	23	44	11	36	33	35
28	34	22	56	55	22	55	33	33	11	14	13	23
29	55	22	35	45	62	44	22	-9	33	44	11	34
30	33	23	24	26	66	55	33	44	11	36	13	66
31	44	23	24	25	43	44	22	44	22	33	13	34
32	11	-9	34	22	44	44	33	44	11	-9	33	23
33	44	33	45	26	22	55	22	55	-9	34	13	24
34	55	-9	24	66	44	55	23	34	22	34	12	44
35	11	22	25	55	22	55	22	33	11	33	13	35
36	11	33	45	35	22	44	22	34	22	24	11	35
37	33	-9	55	44	-9	-9	11	45	11	14	13	33
38	54	23	24	25	-9	55	23	35	11	46	13	44
39	11	22	55	25	62	-9	23	45	11	33	13	35

626

627

628

629

630

631

Tabela S1. Perfil molecular dos 96 cafeiros em estudo. A codificação de cada indivíduo se encontra na Tabela 1. Os 12 *primers* microsatélites utilizados nas análises se encontram em cada coluna. (continua)

40	11	22	66	22	56	45	33	45	11	-9	31	44
41	54	22	45	77	44	44	-9	45	11	33	33	44
42	44	22	25	25	44	56	-9	45	11	34	23	23
43	54	22	22	25	22	56	23	15	11	33	33	44
44	11	22	44	55	22	55	13	33	11	23	11	23
45	11	22	-9	56	44	22	12	33	22	34	11	23
46	11	22	55	22	24	45	23	44	11	-9	13	55
47	45	22	24	25	42	-9	23	35	22	46	13	35
48	-9	22	35	25	64	55	33	45	11	11	33	-9
49	11	33	45	77	24	45	11	35	11	13	11	11
50	11	22	25	-9	22	44	23	35	11	13	31	33
51	44	22	45	25	22	53	23	33	11	13	11	23
52	11	22	44	-9	22	55	22	45	11	23	11	34
53	11	-9	45	55	24	36	-9	35	22	34	22	-9
54	34	33	45	23	36	44	23	35	11	46	11	22
55	44	33	44	23	42	-9	13	55	11	14	22	22
56	34	11	36	22	24	44	33	55	11	16	11	24
57	55	33	45	-9	12	55	13	33	11	46	11	35
58	54	33	45	22	22	55	23	-9	11	34	11	45
59	55	22	44	22	22	55	23	45	11	33	11	55

632

633

634

635

636

637

Tabela S1 . Perfil molecular dos 96 cafeeiros em estudo. A codificação de cada indivíduo se encontra na Tabela 1. Os 12 *primers* microsatélites utilizados nas análises se encontram em cada coluna. (continua)

60	11	22	44	35	22	44	12	34	11	24	11	45
61	11	-9	-9	66	66	55	22	44	22	44	13	45
62	-9	22	55	22	42	65	23	22	22	13	22	44
63	44	33	45	22	44	44	23	35	11	14	11	45
64	11	11	24	-9	44	45	33	45	-9	16	11	45
65	11	33	45	23	44	45	23	53	11	24	11	45
66	11	-9	-9	-9	22	55	22	45	33	34	11	25
67	55	33	25	26	-9	45	23	35	11	36	23	34
68	11	22	-9	35	22	65	13	33	11	44	11	34
69	33	33	34	45	44	44	33	16	22	33	13	34
70	33	22	24	25	42	44	23	43	22	14	33	23
71	44	33	45	25	42	55	22	55	11	33	13	35
72	44	33	45	22	66	55	-9	55	11	-9	33	22
73	44	22	24	22	22	56	33	45	11	36	11	24
74	53	22	25	25	22	55	23	35	11	34	13	34
75	33	33	45	34	22	33	12	43	22	24	13	35
76	11	22	55	33	22	11	22	43	11	24	13	33
77	55	33	45	56	24	55	22	44	22	34	11	33
78	53	33	45	25	42	53	23	44	33	36	13	45
79	55	33	45	26	24	44	11	45	11	36	13	22

638

639

640

641

642

643

Tabela S1. Perfil molecular dos 96 cafeeiros em estudo. A codificação de cada indivíduo se encontra na Tabela 1. Os 12 *primers* microsatélites utilizados nas análises se encontram em cada coluna. (conclusão)

80	45	33	45	45	66	43	33	33	22	44	11	34
81	54	33	45	22	22	44	33	44	11	66	13	44
82	53	22	44	12	22	55	22	35	11	36	11	44
83	33	-9	25	22	22	44	11	44	11	36	13	23
84	11	22	55	33	66	55	22	44	11	34	13	45
85	45	22	25	36	22	65	22	34	22	34	11	34
86	43	22	25	56	42	45	23	44	22	14	13	33
87	11	22	-9	25	44	55	23	35	11	36	13	34
88	11	-9	-9	-9	66	53	33	44	11	56	13	24
89	54	33	44	-9	44	55	13	44	11	36	13	33
90	55	22	45	22	22	44	-9	35	11	34	11	23
91	11	33	25	22	22	45	23	45	11	36	11	33
92	55	33	34	45	46	66	22	34	33	34	13	33
93	11	33	24	-9	22	55	23	33	22	33	33	23
94	45	33	45	22	44	44	22	55	11	34	33	44
95	45	34	45	22	34	56	13	44	11	36	33	34
96	45	33	45	-9	-9	44	33	44	11	11	13	22

Legenda: C02: CaEst 02, C29: CaEst 029, C02: CaEst 02, C29: CaEst 029, C71: Caest 071, C103: CaEst 103.

644

645

646

647

648

649

650