

EDILAINE SILVA ARCANJO

**PAINÉIS DE MARCADORES DE BAIXA DENSIDADE PARA A PREDIÇÃO
GENÔMICA DE *Coffea arabica* L.**

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Estatística Aplicada e Biometria, para obtenção do título de *Magister Scientiae*.

Orientadora: Ana Carolina Campana Nascimento

Coorientadores: Moysés Nascimento
Camila Ferreira Azevedo

**VIÇOSA - MINAS GERAIS
2021**

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

T

A668p
2021
Arcanjo, Edilaine Silva, 1991-
Painéis de marcadores de baixa densidade para predição
genômica de Coffea arábica L. / Edilaine Silva Arcanjo. –
Viçosa, MG, 2021.

1 dissertação eletrônica (41 f.): il.

Orientador: Ana Carolina Campana Nascimento.
Dissertação (mestrado) - Universidade Federal de Viçosa.
Referências bibliográficas: f. 36-41.
DOI: <https://doi.org/10.47328/ufvbbt.2021.103>
Modo de acesso: World Wide Web.

1. Café - Melhoramento genético. 2. Genômica. 3. BLUP.
I. Universidade Federal de Viçosa. Departamento de Estatística.
Programa de Pós-Graduação em Estatística Aplicada e
Biometria. II. Título.

CDD 22. ed. 633.732

Bibliotecário(a) responsável: Alice Regina Pinto CRB6 2523

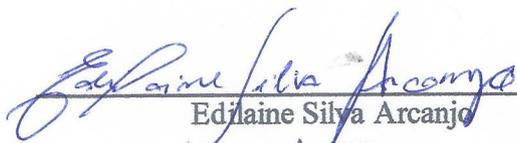
EDILAINÉ SILVA ARCANJO

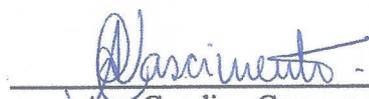
**PAINÉIS DE MARCADORES DE BAIXA DENSIDADE PARA A PREDIÇÃO
GENÔMICA DE *Coffea arabica* L.**

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Estatística Aplicada e Biometria, para obtenção do título de *Magister Scientiae*.

APROVADA: 22 de julho de 2021.

Assentimento:


Edilaine Silva Arcanjo
Autora


Ana Carolina Campana Nascimento
Orientadora

Dedico a Deus, sem Ele nada teria acontecido.
Aos meus amados pais, pelo amor e por sempre acreditarem em mim. À minha irmã Maria Izabel, que sempre foi um exemplo de dedicação e perseverança. E ao meu amado esposo, pela compreensão e companheirismo.

AGRADECIMENTOS

A Deus pelo dom da vida, por ter sido o meu rochedo e meu porto seguro em todos os momentos. Agradeço à intersecção de Nossa Senhora, essa mãe protetora e carinhosa. Agradeço aos santos de devoção e ao meu querido anjo da guarda. Como a fé me fortaleceu nos dias difíceis e não me deixou perder a esperança em dias melhores.

Agradeço a Universidade Federal de Viçosa (UFV) e ao curso de Pós-Graduação em Estatística Aplicada e Biometria pela oportunidade e pelo ensino de excelência.

Aos professores do departamento de estatística, de maneira especial ao José Ivo Ribeiro Júnior que sempre foi um incentivador e amigo.

A minha orientadora Ana Carolina Campana Nascimento, pela dedicação, ensinamentos e conselhos. Uma verdadeira mãe! Em minhas orações sempre agradei à Deus por tê-la colocado em minha vida!

Ao coorientador Moysés Nascimento, pelos conselhos, ensinamentos e por transmitir com leveza e alegria seus conhecimentos.

À coorientadora Camila Ferreira Azevedo, por sua disponibilidade, conselhos, dedicação e por ser um exemplo como professora.

Agradeço aos meus pais, Miguel e Marlene, que são um exemplo para mim, por não terem medido esforços para que eu pudesse estudar. Lembro do sorriso e alegria de minha mãe ao receber a notícia de que eu havia entrado na universidade e posteriormente no mestrado. É lindo ver a satisfação de vocês ao comemorar minhas conquistas! Gratidão!

À minha irmã Maria Izabel, que sempre foi um exemplo de mulher determinada. Por sempre me incentivar a estudar e a lutar pelos meus sonhos.

Ao meu esposo Roberto de Freitas, que sempre me apoiou e me incentivou a lutar por meus sonhos.

Aos amigos Ithalo e Cynthia, por terem ajudado com as rotinas computacionais utilizadas neste estudo

Ao amigo Lucas Silveira, pela sua disponibilidade em me ajudar com os scripts. Que pessoa maravilhosa! Nunca vou me esquecer daquele dia em que pacientemente me auxiliou com as rotinas computacionais, nas quais tinha dificuldade. Deus abençoe imensamente você!

À Gabriela França Oliveira, por ser tão meiga, amiga e por me ajudar sempre!

Aos membros da banca examinadora Isabela de Castro Sant'anna, Eveline Teixeira Caixeta, Camila Azevedo Ferreira e Moysés Nascimento pelo aceite ao convite de participação da banca.

Aos amigos do Laboratório de Inteligência Computacional e Aprendizado Estatístico (LICAE), por me proporcionarem momentos de alegria, conhecimentos e aprendizado de trabalho em equipe.

Aos amigos adquiridos ao longo de minha trajetória na UFV e àqueles do Programa de Pós-Graduação em Estatística Aplicada e Biometria.

À Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) e à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela bolsa de estudo concedida para o desenvolvimento deste trabalho.

“Tudo é possível ao que crê”.

Marcos

9:23

BIOGRAFIA

EDILAINE SILVA ARCANJO, filha de Miguel Arcanjo da Silva e Marlene Silva, nasceu no dia 26 de novembro de 1991, em Viçosa, Estado de Minas Gerais.

Em março de 2014, ingressou no curso de Licenciatura em Matemática na Universidade Federal de Viçosa, Viçosa-MG, graduando-se em janeiro de 2019.

Em agosto de 2019, iniciou o curso de Mestrado no Programa de Pós-Graduação em Estatística Aplicada e Biometria na Universidade Federal de Viçosa, submetendo-se à defesa da dissertação em julho de 2021.

RESUMO

ARCANJO, Edilaine Silva, M.Sc., Universidade Federal de Viçosa, julho de 2021. **Painéis de marcadores de baixa densidade para a predição genômica de *Coffea arabica* L.** Orientadora: Ana Carolina Campana Nascimento. Coorientadores: Moysés Nascimento e Camila Ferreira Azevedo.

Os processos de melhoramento genético são primordiais para o desenvolvimento de novas cultivares. Em decorrência da importância da cafeicultura brasileira, esse setor tem sofrido transformações através das pesquisas em programas de melhoramento. Os progressos do *Coffea arabica* atingidos pelo melhoramento genético têm propiciado a aquisição e recomendação de inúmeras cultivares que possuem características que a elas foram adicionadas por essa técnica. Entretanto, um dos maiores impasses do melhoramento genético vegetal é que para a obtenção de uma nova cultivar, o processo é muitas vezes lento e demorado. Dessa forma, o uso da biotecnologia, com a utilização dos marcadores moleculares, apresentou-se como uma alternativa para amenizar esse problema. Neste contexto, foi proposto a seleção genômica ampla (*Genome Wide Selection*-GWS), que parte do pressuposto que todos os segmentos do genoma colaboram para a variação genética e cada segmento está em alto desequilíbrio de ligação (LD) com no mínimo um marcador genético conhecido. A GWS fundamenta-se nos marcadores moleculares do tipo SNP (*Single Nucleotide Polymorphism*), que são abundantemente distribuídos ao longo do DNA. Com o advento dos SNPs, os valores genéticos genômicos estimados (GEBVs) puderam ser calculados através dos efeitos desses marcadores. Desse modo, os SNPs têm proporcionado a melhor cobertura do genoma; no entanto, normalmente a execução da seleção genômica requer uma grande genotipagem populacional para os indivíduos de treinamento e os candidatos à seleção, o que pode ocasionar em um aumento do custo total do programa de melhoramento. Assim, este trabalho teve por objetivo avaliar a viabilidade do uso de painéis de marcadores de baixa densidade na predição do GEBV de características economicamente importantes de *C. arabica*, com a finalidade de reduzir os custos de genotipagem a partir da utilização de chips customizados. Os resultados obtidos neste estudo demonstraram que o uso desses painéis na GWS pode ser uma ferramenta útil para o melhoramento dessa espécie, uma vez que modelos baseados nestes painéis apresentaram boas estimativas de capacidades preditivas e substanciais valores de concordância em termos de seleção quando comparados à modelos de maior densidade de marcadores.

Palavras-chave: Melhoramento Genético. Café. Seleção Genômica. G-BLUP.

ABSTRACT

ARCANJO, Edilaine Silva, M.Sc., Universidade Federal de Viçosa, July, 2021. **Low density marker panels for the genomic prediction of *Coffea arabica* L.** Advisor: Ana Carolina Campana Nascimento. Co-advisors: Camila Ferreira Azevedo and Moysés Nascimento.

The processes of genetic improvement are paramount for the development of new cultivars. Due to the importance of Brazilian coffee, this sector has undergone transformations through research in breeding programs. The progress of *Coffea arabica* achieved by genetic improvement has provided the acquisition and recommendation of numerous cultivars that have characteristics that have been added to them by this technique. However, one of the biggest impediments of plant genetic improvement is that to obtain a new cultivar, the process is often slow and time-consuming. Thus, the use of biotechnology, with the use of molecular markers, was presented as an alternative to soften this problem. In this context, broad genomic selection (Genomic Wide Selection-GWS) was proposed, which assumes that all segments of the genome contribute to genetic variation and each segment is in linkage disequilibrium (LD) with at least one known genetic marker. GWS is based on molecular markers of the SNP (*Single Nucleotide Polymorphism*) type, which are abundantly distributed along the DNA. With the advent of SNPs, the estimated genomic genetic values (GEBVs) could be calculated through the effects of these markers. In this way, SNPs has provided the best coverage of the genome; however, usually the execution of genomic selection requires a large population genotyping for training individuals and candidates for selection, which can lead to an increase in the total cost of the breeding program. Thus, this work aims to evaluate the feasibility of using panels of low-density markers in predicting the GEBV of economically important characteristics of *C. arabica* in order to reduce genotyping costs from the use of customized chips. The results obtained in this study demonstrated that the use of these low density panels in GWS can be a useful tool for the improvement of this species, since models based on these panels presented good estimates of predictive capacities and substantial values of agreement in terms of selection when compared to models with higher density of markers.

Keywords: Genetic Improvement. *Coffea*. Genomic Selection. G-BLUP.

SUMÁRIO

1. INTRODUÇÃO.....	11
2. REFERENCIAL TEÓRICO.....	13
2.1 Melhoramento do Café.....	13
2.2 Seleção Genômica	14
2.3 Método G-BLUP	16
2.4 Capacidade Preditiva e Herdabilidade Genômica	18
2.5 Validação Cruzada.....	18
2.6 Coeficiente Cohen's Kappa.....	19
3. MATERIAL E MÉTODOS.....	20
3.1 Condução Experimental	20
3.2 Material Genético	20
3.3 Avaliações Fenotípicas	21
3.4 Análise de qualidade dos marcadores SNP	22
3.5 Análises Fenotípicas	22
3.6 Predição através do método G-BLUP	23
3.7 Densidade de Marcadores.....	24
3.8 Intersecção e União dos SNPs.....	25
4. RESULTADOS E DISCUSSÃO.....	26
5. CONCLUSÃO.....	35
6. REFERÊNCIAS BIBLIOGRÁFICAS	36

1. INTRODUÇÃO

O agronegócio é um dos ramos que movimenta mensalmente bilhões de reais na receita brasileira, contribuindo de modo substancial para a redução do déficit e para o superávit na balança comercial (FIESP, 2021). Dentre os principais produtos do agronegócio voltados para a exportação, o café ocupa a 5ª posição desse ranking, sendo o Brasil o maior produtor, exportador e o segundo maior consumidor da bebida mundialmente (MAPA, 2017; FOODCHAIN, 2018).

O gênero *Coffea* totaliza 104 espécies (SAKIYAMA et al., 2015), sendo o *Coffea arábica* L. e *Coffea canephora* as espécies economicamente mais relevantes (BESSADA et al., 2018). No Brasil, o *C.arábica* é produzido nas regiões Sul, Sudeste, Centro-Oeste e Nordeste (SUMÁRIO EXECUTIVO DO CAFÉ, 2020). Em setembro de 2020, as exportações dos cafés do Brasil alcançaram o maior volume registrado para o mês, somando 3,8 milhões de sacas de 60 kg, com o *C.arábica* contribuindo com 74,8% do volume da embarcação, ou seja, 2,8 milhões de sacas, contra 17,7% de participação do café *Kouillou* (Conilon) que foi responsável pela embarcação de 672,5 mil sacas (CECAFÉ, 2020).

Em decorrência da importância da cafeicultura brasileira, esse setor vem sofrendo transformações ao longo do tempo através das pesquisas em programas de melhoramento genético. Especificamente, tais programas tem por objetivo desenvolver novas cultivares que se adaptem às distintas condições de cultivo, aumentem a produção e que sejam capazes de trazer desenvolvimento socioeconômico das regiões de cultivo (FREITAS et al., 2007). Os progressos do cafeeiro arábica atingidos pelo melhoramento genético têm propiciado a aquisição e recomendação de inúmeras cultivares que possuem características que a elas foram adicionadas pelo processo de melhoramento, tais como: maturação uniforme; resistência ou tolerância à pragas e doenças; produtividade elevada, estável e com precocidade na primeira colheita; tolerância à geada, à seca e ao alumínio tóxico; arquitetura adequada para adensamento; porte baixo e formato de copa adequado para colheita mecanizada; e elevado tamanho de grão e qualidade de bebida (CARVALHO, 2008).

Entretanto, um dos maiores impasses no processo de melhoramento genético vegetal, especificamente em espécies perenes, é que para a obtenção de uma nova cultivar, o processo é muitas vezes lento e demorado. Assim, o uso da biotecnologia, com a utilização dos marcadores moleculares, apresentou-se como uma alternativa para amenizar esse problema. Os marcadores de DNA relacionados à genes de interesse são primordiais na seleção de genótipos superiores e constituem uma importante ferramenta que auxilia nos processos de seleção

(CAIXETA et al., 2015). Neste contexto, destaca-se a utilização dos marcadores do tipo SNP (*Single Nucleotide Polymorphism*), que se baseiam na identificação de polimorfismos decorrentes de uma única base no genoma (RESENDE et al., 2012).

Meuwissen et al. (2001), propuseram a Seleção Genômica Ampla (GWS), que parte do pressuposto que todos os segmentos do genoma colaboram para a variação genética e cada segmento está em alto LD (*Linkage Disequilibrium*) com no mínimo um marcador genético conhecido. Dessa forma, os efeitos dos marcadores genéticos, como por exemplo, os SNPs, em todo o genoma podem ser estimados e usados para prever o mérito genético de candidatos à seleção (MEUWISSEN et al., 2001). Assim, a GWS tem se mostrado uma ferramenta valiosa para estimar os valores genéticos na criação de plantas e animais (ZHANG et al., 2019).

Os métodos estatísticos utilizados na GWS podem ser categorizados em métodos baseados em parentesco e métodos baseados nos efeitos dos marcadores. Os métodos baseados em parentesco predizem de forma direta os valores genéticos usando informações de parentesco sem a necessidade de estimar os efeitos dos marcadores (ZHANG et al., 2019). Dentre esses métodos, o G-BLUP (*Genomic Best Linear Unbiased Prediction*), que utiliza uma matriz de parentesco com base em informações genômicas que gera coeficientes de parentescos mais acurados entre os indivíduos, ampliou a acurácia da estimativa dos valores genéticos e tem sido muito utilizado no melhoramento de plantas e animais (VANRADEN, 2008).

Por outro lado, os métodos baseados no efeito dos marcadores possuem duas etapas: (1) estimam os efeitos dos marcadores genéticos e posteriormente (2) acumulam os efeitos para obter os valores genéticos genômicos estimados (GEBV) de cada indivíduo. Eles são representados pelos métodos do alfabeto Bayesiano e o método de regressão de cumeeira RR-BLUP (*Ridge Regression-Best Linear Unbiased Prediction*) (WHITTAKER et al., 2000; ENDELMAN, 2011).

Os chips SNP de alta densidade têm proporcionado a melhor cobertura do genoma; entretanto, normalmente a execução da seleção genômica requer uma grande genotipagem populacional para os indivíduos de treinamento e os candidatos à seleção, o que pode ocasionar em um aumento do custo total do programa de melhoramento (BANDEIRA e SOUSA et al., 2019). Uma maneira de minimizar estes custos, seria a genotipagem dos candidatos à seleção com um painel de SNPs com densidade reduzida. Estudos em dados de arroz e milho (BANDEIRA e SOUSA et al., 2019), aquicultura (KRIARIDOU et al., 2020), gado leiteiro (ALILOO et al., 2018), suínos (WELLMANN et al., 2013) dentre outros, mostraram que através de um painel de SNPs de baixa densidade é possível selecionar os marcadores mais

informativos para melhorar a acurácia e construir um chip SNP de custo mais acessível para implementar a seleção genômica em programas de melhoramento.

Diante do exposto, neste trabalho objetivou-se avaliar a viabilidade do uso de painéis de marcadores de baixa densidade na predição do valor genético genômico de características economicamente importantes de *C.arábica*, visando obter um modelo de predição mais parcimonioso, com a finalidade de reduzir os custos de genotipagem nos programas de melhoramento dessa espécie de cafeeiro.

2. REFERENCIAL TEÓRICO

2.1 Melhoramento do Café

Os processos de melhoramento genético são primordiais para o desenvolvimento de novas cultivares. A princípio, o melhoramento genético de uma determinada cultura, está diretamente relacionado a um processo de modificação das plantas com o objetivo de aumentar a produtividade. Com o surgimento da genética, no século XIX e o sequente desenvolvimento dos métodos de seleção com avaliações biométricas, novas técnicas foram acrescentadas, tendo em vista a resistência de pragas e doenças que minimizavam de modo direto a produção (FILHO; BORDIGNON, 2008). Mais tarde, também foi dado destaque à tolerância a estresses ambientais, como uma estratégia para minimizar os riscos relacionados as perdas de produção associados às causas aleatórias como frio, seca, vento, etc.

O café é um dos principais produtos de exportação do Brasil. As duas espécies de cafeeiro mais cultivadas no país são *Coffea arábica* L. (café arábica) e *Coffea canephora* Pierre (café robusta) (RESENTE et al., 2001). A edição do relatório mensal do Conselho dos Exportadores de Café do Brasil (Cecafé) em setembro de 2020, destacou que nesse mês o Brasil exportou um total de 3,8 milhões de sacas de café, sendo 74,8% do volume total correspondente ao café arábica, contra 17,7% do café conilon (robusta) (CECAFÉ, 2020). Segundo Resende et al. (2001), em função da relevância da cultura do café para o Brasil, os programas de melhoramento do cafeeiro promovidos no país desempenham importância fundamental como fator de incremento da produção, produtividade e de avanço socioeconômico das regiões de cultivo.

A criação de cultivares resistentes às pragas e patologias possuem um importante papel para o estabelecimento da produtividade e na redução dos custos de produção (FREITAS et al., 2007). Os fungos *Hemileia vastatrix* e *Colletotrichum kahawae* causadores da ferrugem-alaranjada e da antracnose dos frutos (Coffee Berry Disease-CBD) respectivamente, são

doenças que afetam diretamente o cultivo e a produção de espécies perenes como o *C. arábica*. Já o bicho mineiro representa a principal praga que afeta os cafezais (FILHO; BORDIGNON, 2008). Nesse sentido, o uso de cultivares resistentes vem como opção de controle eficaz e econômico, além de prevenir a degradação do ambiente e a contaminação dos operários rurais por fungicidas (MAIA et al., 2016; TALHINHAS et al., 2017; ZAMBOLIM, 2016).

Em contrapartida, somente resistências às doenças e pragas não é o bastante para uma nova cultivar ser aceita pelos produtores, também é necessário possuir vigor vegetativo, alto potencial produtivo, porte e arquitetura das plantas, tamanhos dos grãos, qualidade da bebida, entre outras características de interesse econômico (PEREIRA e OLIVEIRA, 2015; VAN DER VOSSSEN et al., 2015). Desse modo, o processo de melhoramento do café é complexo e demorado, uma vez que por ser uma espécie vegetal perene, apresenta alguns aspectos biológicos particulares, tais como ciclo reprodutivo longo, sobreposição de gerações, bienalidade produtiva, expressão dos caracteres ao longo de vários anos e diferenças em precocidade e longevidade produtiva (SERA, 2001). Diante desses impasses agrônômicos, no processo de melhoramento genético do cafeeiro, torna-se necessário a utilização de técnicas que possam aumentar a acurácia da seleção de genótipos superiores, sendo sugerido a utilização de métodos especiais para a seleção e predição dos valores genéticos (OLIVEIRA et al., 2011; PETEK et al., 2009).

2.2 Seleção Genômica

Para o melhoramento genético ser efetivo, o geneticista precisa desenvolver e identificar genótipos superiores (RESENDE et al., 2008). Nessas ações, a seleção desempenha um papel primordial, uma vez que ela define os cruzamentos que serão realizados com o propósito de desenvolver novos genótipos e na prescrição dos indivíduos superiores que serão utilizados comercialmente (RESENDE et al., 2012). Dessa forma, o incentivo da genética molecular em benefício do melhoramento genético aplicado é o uso direto das informações do DNA na seleção que proporciona alta eficiência seletiva e grande rapidez na aquisição de ganhos genéticos (RESENDE et al., 2012). Visando esses propósitos, Meuwissen et al. (2001), propuseram o método de seleção intitulado seleção genômica ou seleção genômica ampla (*Genome Wide Selection-GWS*), que pode ser empregado em todas as famílias em avaliação dos programas de melhoramento (RESENDE et al. 2012).

A GWS fundamenta-se nos marcadores genéticos moleculares do tipo SNP que são abundantemente distribuídos através do genoma e são escolhidos em detrimento a outros

marcadores por serem codominantes, possuírem baixa taxa de mutação e por permitirem genotipagem em larga escala (CAETANO, 2009; RESENDE et al. 2012). Ademais, os efeitos genéticos dos marcadores SNP, em desequilíbrio de ligação com os QTLs, podem ser usados como base para reconhecimento de indivíduos candidatos à seleção, o que ampliaria a acurácia na avaliação genética (GODDARD; HAYES, 2007; RESENDE et al., 2012). Desse modo, as predições derivadas de dados fenotípicos e de marcadores SNP em alta densidade em uma geração são então usadas para obtenção dos valores genéticos genômicos (GEBV) dos indivíduos de qualquer geração subsequente, tendo por base os seus próprios genótipos marcadores (RESENDE et al. 2012).

Na GWS são utilizadas três populações: população de estimação (onde são avaliados os fenótipos e genótipos para estimação dos efeitos dos marcadores), população de validação (realizada com indivíduos fenotipados e genotipados, que podem ser diferentes ou não da população de estimação, no entanto, utilizando-se os efeitos pré-estimados dos marcadores para prever os fenótipos) e a população de seleção (em que não há necessidade de fenotipar os indivíduos e as equações de predição derivadas na população de estimação são utilizadas para prever os fenótipos futuros) (GODDARD; HAYES, 2007; GRANATO et al., 2013).

No contexto da seleção assistida por marcadores e da predição genômica, um bom método de estimação é aquele que dentre outras propriedades, efetua a regularização do processo de estimação devido ao elevado número de marcadores, utilizando para esse caso estimadores do tipo *shrinkage*. Esse tipo de estimadores, lidam com o principal problema da GWS, que consiste na estimação de uma considerável quantidade de efeitos a partir de uma limitada quantidade de observações, bem como as colinearidades provenientes do desequilíbrio de ligação entre marcadores (RESENDE et al., 2008).

Os principais métodos para a GWS podem ser divididos em três grandes classes: regressão explícita, implícita e com redução de dimensionalidade (RESENDE et al., 2012). A primeira classe pode ser dividida em dois subgrupos: métodos de regressão linear, que incluem o RR-BLUP (*Random Regression Best Linear Unbiased Predictor*), LASSO (*Least Absolute and Shrinkage and Selection Operator*), Rede Elástica (*Elastic Net- EN*) e RR-BLUP-Het. Esses métodos, também são chamados de métodos de estimação penalizada e a diferença entre eles consiste nas distintas funções de penalização utilizadas. O segundo subgrupo, engloba os métodos de regressão não-linear (métodos bayesianos), dentre eles, o Bayes A, Bayes B, Bayes C Π , entre outros. Esses métodos podem ser melhores quando os efeitos dos QTLs não são

normalmente distribuídos, devido a presença de genes de efeitos maiores. Segundo Resende et al. (2012), na estimação bayesiana, o encurtamento das estimativas dos efeitos do modelo é controlado pela distribuição *a priori* assumida para esses efeitos. Assim, distintas distribuições *a priori* conduzem a diferentes encurtamentos.

Os métodos de Redes Neurais, Regressão Kernel e RKHS (*Reproducing Kernel Hilbert Spaces*), são da classe de regressão implícita. Ademais, como exemplo de métodos de redução de dimensionalidade, podemos citar os de Regressão via Componentes Principais (PCR), Regressão via Componentes Independentes (ICR) e Regressão via Quadrados Mínimos Parciais (PLS) (RESENDE et al., 2012). A seguir apresentaremos o método G-BLUP que será utilizado no presente estudo.

2.3 Método G-BLUP

A classe dos modelos de seleção do tipo misto apresentado por Henderson (1973), considera o método BLUP (melhor preditor linear não viesado), que possibilitou que grandes quantidades de ganho genético fossem alcançadas em muitos programas de melhoramento. Segundo Resende et al. (2012), o procedimento BLUP pode ser definido da seguinte forma:

- B: reduz a variância do erro de predição (PEV), isto é, aumenta a precisão.
- L: constitui uma função linear das observações.
- U: é não viesado, essa propriedade juntamente com a redução da PEV, aumenta a acurácia na classe dos preditores não viciados.
- P: preditor de uma variável aleatória.

Dessa forma, temos duas derivações da metodologia BLUP, a “tradicional” que se baseia em informações de *pedigree* e que foi amplamente utilizada no melhoramento animal. Entretanto, suas formas de reprodução limitavam o uso de informações de *pedigree* no melhoramento de plantas (HIDALGO et al., 2015; HICKEY et al., 2017). A segunda metodologia utilizada para combinar informações genômicas em BLUP foi introduzida por VanRaden (2008) e Harbier et al., (2007) que é o *Genomic Best Linear Unbiased Prediction* (G-BLUP) que faz uso de relações genômicas para estimar o mérito genético de um indivíduo.

O G-BLUP utiliza uma matriz de parentesco genômico (MPG) que fornece coeficientes de parentesco mais precisos entre os indivíduos, possibilitando o aumento da acurácia da estimativa dos valores genéticos (ZHANG et al., 2019). Villaneuva et al., (2005) avaliaram o uso da MPG como um modo de avaliação genômica e propuseram que quando a

variação genética é explicada por muitos locos de características quantitativas (QTLs) de pequeno efeito, o BLUP usando uma MPG pode ser utilizado para produzir estimativas de maior acurácia do que o BLUP baseado em *pedigree* por apresentar relações aditivas entre indivíduos baseadas em informações usando marcadores de DNA compartilhados (CLARK; WERF, 2013).

Ademais, o G-BLUP é um método de regularização que faz uso de estimadores do tipo *shrinkage*, que assumem os efeitos de marcadores como variáveis de efeito aleatório. Esses estimadores lidam de modo adequado com o principal impasse da GWS que consiste na estimação de um grande número de efeitos com base em um limitado número de informações e também as colinearidades provenientes do desequilíbrio de ligação entre os marcadores (RESENDE et al., 2008). Segundo Resende et al. (2012), o modelo linear misto geral para estimar os efeitos genéticos aditivos a partir das informações fenotípicas pode ser definido por:

$$y = Xb + Zg + e, \quad (1)$$

em que:

y é um vetor de observações fenotípicas;

b é um vetor de efeitos fixos;

g é um vetor de efeitos aditivos individuais aleatórios;

X e Z são as matrizes de incidência dos efeitos fixos e aleatórios, respectivamente;

e é um vetor de erros normais aleatórios com $e \sim N(0, I\sigma_e^2)$ e σ_e^2 é variância residual. Para estimar os efeitos dos marcadores, é ajustado o modelo equivalente a (1), dado segundo Resende et al. (2012) por:

$$y = Xb + ZWm + e, \quad (2)$$

em que:

y é um vetor de observações fenotípicas;

b é um vetor de efeitos fixos;

m é um vetor de efeitos aleatórios dos marcadores;

X é uma matriz de incidência dos efeitos fixos;

W é uma matriz de incidência de m com $g = Wm$.

As equações de modelo misto para a predição de g por meio do método G-BLUP, são dadas por:

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + G^{-1} \frac{\sigma_e^2}{\sigma_g^2} \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{g} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \end{bmatrix},$$

em que $G = (WW')/k$ é a matriz de parentesco genômica aditiva, sendo $k = 2 \sum_i^n p_i (1 - p_i)$ o parâmetro de escala e σ_g^2 é o componente de variação genética total. O valor genético genômico global do indivíduo j é dado por $VGG = \hat{g}_j = \sum_i w_{ij} \hat{m}_i$. Dessa forma, com base na estimação dos valores genéticos (\hat{g}_j) pelo G-BLUP, as estimativas dos efeitos dos marcadores (\hat{m}) podem ser obtidas através de: $\hat{g} = W\hat{m} \Rightarrow \hat{m} = (W'W)^{-1}W'\hat{g}$.

A análise feita pelo G-BLUP é computacionalmente eficiente, uma vez que requer a resolução de um menor número de equações (RESENDE et al., 2012), sendo capaz de atender aos requisitos de predição de características com arquitetura genética complexa, como no caso de *C.arábica*.

2.4 Capacidade Preditiva e Herdabilidade Genômica

A capacidade preditiva ($r_{y\hat{y}}$) é uma medida que indica o quanto um método é capaz de fazer predições acuradas. Ela é obtida por meio da correlação entre os valores fenotípicos observados (y) e os valores genômicos estimados (\hat{y}), isto é, $r_{y\hat{y}} = cor(y, \hat{y})$ (RESENDE et al., 2012).

Já a herdabilidade (h^2) é o coeficiente que expressa o quanto da variabilidade fenotípica de uma característica é de origem genética, ou seja, $h^2 = \frac{\sigma_g^2}{\sigma_f^2}$, em que σ_g^2 é a variância genotípica e σ_f^2 é a variância fenotípica (RESENDE et al., 2014).

2.5 Validação Cruzada

De acordo com Resende et al. (2012), a população em estudo pode ser dividida em dois subconjuntos, o primeiro denominado população de estimação ou treinamento e o segundo denominado população de validação. Na população de estimação, estão os indivíduos com informações fenotípicas e genotípicas, esta população é utilizada com a finalidade de prever os efeitos dos marcadores sobre a característica de interesse. Já na população de validação, também se encontra indivíduos com informações fenotípicas e genotípicas, porém, com base

nos efeitos dos marcadores preditos na população de treinamento, é possível estimar os valores genéticos genômicos dos indivíduos desta população.

Considerando as duas populações descritas anteriormente, uma das formas de validação muito utilizada é a chamada validação cruzada. Esta metodologia é empregada para avaliar a capacidade de generalização de um modelo preditivo, a partir de um conjunto de dados (KOHAVI, 1995). Na técnica de validação cruzada denominada *k-fold*, o conjunto de dados composto por n elementos é dividido em k grupos com m elementos cada, tal que $n = mk$. Posteriormente, um grupo é utilizado como população de validação e os $(m - 1)$ restantes como população de estimação. Neste processo, todos os grupos em uma das etapas será utilizado como população de validação. A cada ciclo são calculadas as medidas relacionadas à qualidade do modelo ajustado, como a acurácia, capacidade preditiva e o viés.

O uso da validação cruzada, implica em predizer medidas que avaliam o modelo em uma população diferente daquela utilizada para a obtenção das estimativas dos parâmetros. Este procedimento evita a superestimação destas medidas, que ocorre quando o modelo está sendo ajustado e validado com os mesmos indivíduos (TEIXEIRA, 2015).

2.6 Coeficiente Cohen's Kappa

Proposto por Cohen (1960), o coeficiente Cohen's Kappa foi desenvolvido com a finalidade de verificar a concordância entre duas avaliações independentes. Este coeficiente leva em consideração a probabilidade de a concordância ter ocorrido aleatoriamente, o que o torna um índice com maior precisão. A estimativa desse coeficiente é dada a partir da seguinte expressão:

$$\hat{k} = \frac{Pr(a) - Pr(e)}{1 - Pr(e)},$$

em que $Pr(a) - Pr(e)$ é a proporção de observações na qual a concordância ocorreu além do que se esperava ao acaso e $1 - Pr(e)$, representa a proporção de observações que não houve concordância. A medida \hat{k} varia de 0 até 1 e quanto maior o valor de \hat{k} , mais os grupos estão em concordância. Na Tabela 1, está a classificação do índice Cohen's Kappa segundo Landis e Koch (1977). De acordo com estes autores, esse índice é considerado satisfatório quando assume valores de \hat{k} maiores do que 0,4.

Tabela 1: Classificação do índice Cohen's Kappa segundo Landis e Koch (1977).

Índice de Kappa	Classificação
$\hat{k} \leq 0,2$	Ruim
$0,2 < \hat{k} \leq 0,4$	Razoável
$0,4 < \hat{k} \leq 0,6$	Bom
$0,6 < \hat{k} \leq 0,8$	Muito Bom
$0,8 < \hat{k} \leq 1,0$	Excelente

3. MATERIAL E MÉTODOS

3.1 Condução Experimental

O plantio do experimento foi conduzido na estação experimental do Departamento de Fitopatologia da Universidade Federal de Viçosa (UFV) em fevereiro de 2011. A avaliação do experimento foi feita nos anos de 2014, 2015 e 2016. Na área do delineamento, a calagem do solo e a adubação do plantio foram feitas de acordo com as exigências da cultura. As plantas foram arranjadas com espaçamentos de 0,7m entre plantas e 3,0m entre linhas. Não foi utilizado nenhum método de controle fitossanitário contra ferrugem, cercosporiose e bicho-mineiro. Maiores detalhes sobre a condução experimental podem ser consultados em Sousa et al. (2019).

3.2 Material Genético

Para o material genético, foram selecionados para o cruzamento, seis genótipos de dois grupos contrastantes relacionado à resistência à ferrugem do cafeeiro, sendo três genitores do grupo de cultivares Catuaí suscetíveis à *Hemileia vastatrix* e três do grupo Híbrido de Timor (HdT) resistentes à *Hemileia vastatrix*. Desses cruzamentos, resultaram 13 progênes de *C. arábica* do programa de melhoramento da Epamig/UFV/Embrapa com 15 genótipos cada, originando 195 indivíduos, correspondentes às gerações de retrocruzamentos resistentes (RTCr), retrocruzamentos suscetíveis (RTC_s) e F_2 (Tabela 2).

Tabela 2: Progênes de *C. arábica* do programa de melhoramento da Epamig/UFV/Embrapa.

Progênes	Genótipos	Genitor 1	Genitor 2
RTCr1	1-15	H 419-1 c-17	UFV 445-46
RTC _s 2	16-30	H 419-1 c-17	UFV 2143-235
RTCr3	31-45	H 514-8 c-387	UFV 440-10

RTCs4	46-60	H 514-8 c-387	UFV 2154-344
RTCr5	61-75	H 514-7 c-364	UFV 440-10
RTCs6	76-90	H 514-7 c-364	UFV 2154-344
RTCr7	91-105	H 419-10 c-214	UFV 445-46
RTCs8	106-120	H 419-10 c-214	UFV 2143-235
RTCs9	121-135	UFV 2148-57	H 513-5 c-14
F2 10	136-150	H 514-8 c-387	-
F2 11	151-165	H 514-7 c-364	-
F2 12	166-180	H 419-10 c-214	-
F2 13	181-195	H 513-5 c-14	-

3.3 Avaliações Fenotípicas

Neste estudo foram avaliadas oito características fenotípicas de importância agrônoma, sendo elas: comprimento de um ramo plagiotrópico (CRP), número de nós vegetativos (NNV), número de frutos por ramo plagiotrópico (NFRP), diâmetro da copa (Dco), tamanho dos frutos maduros (TFr), incidência de cercosporiose (Cer), infestação de bicho mineiro (BM) e vigor vegetativo (Vig).

Dentre essas características, 5 são categóricas e 2 são contínuas. Para avaliar as características categóricas, foram utilizadas escalas de pontuação. Para a característica TFr em amadurecimento foi utilizada uma escala de pontuação de 1 a 3, sendo 1 para frutos pequenos, 2 para médios e 3 para frutos grandes. Para as características Cer, BM e Vig a avaliação foi através de uma escala com níveis de 1 a 5, em que 1 equivale a genótipos sem sintomas e 5 a genótipos altamente suscetíveis. Para a característica Vig a escala de pontuação variou de 1 a 10, em que 1 corresponde a plantas totalmente depauperadas e 10 a plantas com vigor vegetativo máximo. As características contínuas foram avaliadas como descrito na Tabela 3.

Tabela 3: Descrição das avaliações fenotípicas das características contínuas realizadas em 2014, 2015 e 2016.

Características	Tipo de Avaliação
-----------------	-------------------

Comprimento de um ramo plagiotrópico (CRP) (cm)	Medidas feitas no ramo plagiotrópico localizado no terço médio da planta.
Número de nós vegetativos (NNV)	
Número total de frutos (NFRP)	
Diâmetro da copa (Dco) (cm)	Medida feita na direção transversal à linha de plantio, medindo-se a maior projeção da “saia” do cafeeiro.

3.4 Análise de qualidade dos marcadores SNP

Com base na construção de 40000 sondas polimórficas (Resende et al., 2016), os 195 indivíduos foram genotipados com 21211 SNPs. Esse grupo de marcadores foi submetido a análise de qualidade. Após essa etapa, foram removidos os SNPs que não atenderam aos critérios mínimos especificados pelo MAF (*Minor Allele Frequency*) e CR (*Call Rate*) que, para este estudo, deveriam ser iguais ou superiores a 5% e a 90%, respectivamente. A equação $MAF = 1/\sqrt{2n}$ foi utilizada para obter o nível crítico para o parâmetro MAF, em que n é o número de indivíduos avaliados. Ademais, para a não ocorrência de falsos SNPs (Vidal et al., 2010), foram eliminados àqueles que não apresentaram variância entre os indivíduos pertencentes à população em estudo. As análises de qualidade foram desenvolvidas no software Rbio (BHERING, 2017).

A partir da análise de qualidade, houve uma redução de 3,46% do total de marcadores SNPs, ou seja, dos 21211 marcadores, restaram 20477 SNPs para a realização da GWS.

3.5 Análises Fenotípicas

Os dados fenotípicos foram corrigidos para efeitos de anos, parcelas e interação *anos* × *parcelas* para estimar as capacidades preditivas ($r_{y\hat{y}}$) e as herdabilidades genômicas (h^2) das oito características avaliadas. As análises foram desenvolvidas levando em conta os modelos lineares mistos (procedimento REML/BLUP), através do software Selegen-REML/BLUP (RESENDE, 2016). Foi realizada análise individual das oito características agrônomicas para a estimação dos parâmetros genéticos, para isso, foi ajustado o seguinte modelo estatístico:

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

em que:

y é referente ao vetor de dados;
 u equivale à média geral correspondente a cada ano de avaliação;
 g é referente ao vetor de efeitos de progênies (efeito aleatório);
 p equivale a variância permanente entre plantas (efeito aleatório);
 r corresponde a variância entre tipos de populações (efeito aleatório);
 b corresponde a variância entre parcela (efeito aleatório);
 i equivale a variância da interação progênies x anos (efeito aleatório) e
 e é o vetor de erros (efeito aleatório).

As letras maiúsculas do modelo correspondem às matrizes de incidência para os referidos efeitos.

3.6 Predição através do método G-BLUP

As análises de GWS foram realizadas utilizando o método G-BLUP. Foi ajustado o modelo linear misto geral (RESENDE et al., 2012) para estimar os efeitos genéticos aditivos a partir das informações fenotípicas dado por:

$$y = Xb + Zg + e, \quad (1)$$

em que:

y é um vetor de observações fenotípicas corrigidas para efeitos de anos, parcelas e interação anos x parcelas;

b é um vetor de efeitos fixos;

g é um vetor de efeitos aditivos individuais aleatórios;

X e Z são as matrizes de incidência dos efeitos fixos e aleatórios, respectivamente;

e é um vetor de erros normais aleatórios com $e \sim N(0, I\sigma_e^2)$ e σ_e^2 é variância residual. Para estimar os efeitos dos marcadores, foi ajustado o modelo equivalente a (1), dado segundo Resende et al. (2012) por:

$$y = Xb + ZWm + e, \quad (2)$$

em que:

y é um vetor de observações fenotípicas corrigidas para efeitos de anos, parcelas e interação anos x parcelas;

b é um vetor de efeitos fixos;

X é a matriz de incidência dos efeitos fixos;

m é um vetor de efeitos aleatórios dos marcadores;

W é uma matriz de incidência de m com $g = Wm$.

Para a predição de g por meio do método G-BLUP, foram utilizadas as equações de modelos mistos, dadas por:

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + G^{-1} \frac{\sigma_e^2}{\sigma_g^2} \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{g} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \end{bmatrix},$$

em que $G = (WW')/k$ é a matriz de parentesco genômica aditiva, sendo $k = 2 \sum_i^n p_i (1 - p_i)$ o parâmetro de escala e σ_g^2 é o componente de variação genética total. Assim, o valor genético genômico global do indivíduo j foi dado por $VGG = \hat{g}_j = \sum_i w_{ij} \hat{m}_i$. Dessa forma, com base na estimação dos valores genéticos (\hat{g}_j) pelo G-BLUP, as estimativas dos efeitos dos marcadores (\hat{m}) puderam ser obtidas através de: $\hat{g} = W\hat{m} \Rightarrow \hat{m} = (W'W)^{-1}W'\hat{g}$.

3.7 Densidade de Marcadores

Diferentes densidades de marcadores SNP foram utilizadas para avaliar a viabilidade do uso de painéis de marcadores de baixa densidade na predição do valor genético genômico das 8 características analisadas. Essa viabilidade foi verificada por meio do efeito do número de marcas na capacidade preditiva utilizando-se o método G-BLUP. Além disso, para verificar essa viabilidade, também foram calculadas as concordâncias entre indivíduos selecionados por meio do valor fenotípico e do GEBV predito e as concordâncias entre os indivíduos selecionados por meio do GEBV predito utilizando todos os 20477 marcadores e aqueles selecionados a partir de modelos ajustados com as diferentes densidades.

Neste trabalho, a população composta por 195 cafeeiros, foi dividida em $k = 5$ folds, sendo 156 indivíduos utilizados para estimação/treinamento dos modelos preditivos e 39 indivíduos utilizados para a validação dos mesmos. Foram ajustados inicialmente, para cada característica de interesse, modelos com diferentes densidades de marcadores, isto é: 2047, 4095, 6143, 8191, 10238, 12286, 14334, 16382, 18429 e 20477 SNPs, selecionados com base na magnitude de seus efeitos (maiores efeitos em valor absoluto). Essas densidades correspondem a 10%, 20%, ... e a 100% dos 20477 SNPs, respectivamente.

O processo foi repetido 5 vezes, de modo que cada grupo foi utilizado uma vez como conjunto de validação. Ao final, foi estimada a capacidade preditiva (r_{yy}) e a herdabilidade (h^2), a fim de identificar os subconjuntos de marcadores que apresentaram os melhores valores médios desses parâmetros. Posteriormente, avaliou-se as concordâncias entre indivíduos selecionados por meio do valor fenotípico e do GEBV predito pelos diferentes modelos e através dos indivíduos selecionados por meio do GEBV utilizando todos os 20477 marcadores e àqueles selecionados a partir de modelos ajustados com as diferentes densidades.

A partir dos resultados das primeiras análises, foi escolhido o modelo que obteve as melhores estimativas de capacidade preditiva e as melhores estimativas de concordância.

3.8 Intersecção e União dos SNPs

Com base na escolha do melhor modelo a partir das análises iniciais, buscou-se avaliar diferentes densidades, correspondentes à intersecção e à união dos SNPs referentes ao modelo selecionado dentro de três grupos formados a partir das características avaliadas neste estudo. Os grupos das características foram formados a partir das relações de proximidade entre elas. O conjunto intersecção dos SNPs, é resultante daqueles marcadores comuns às características dentro do seu referente grupo. Já o conjunto união dos SNPs, é resultante dos marcadores que se mostraram importantes à pelo menos uma das características do grupo. A partir dessas etapas, foram obtidas novas densidades, nas quais foram repetidas as análises de capacidade preditiva, herdabilidade e concordâncias.

Todas as análises foram feitas no software R versão 3.6.1 (R CORE TEAM, 2019).

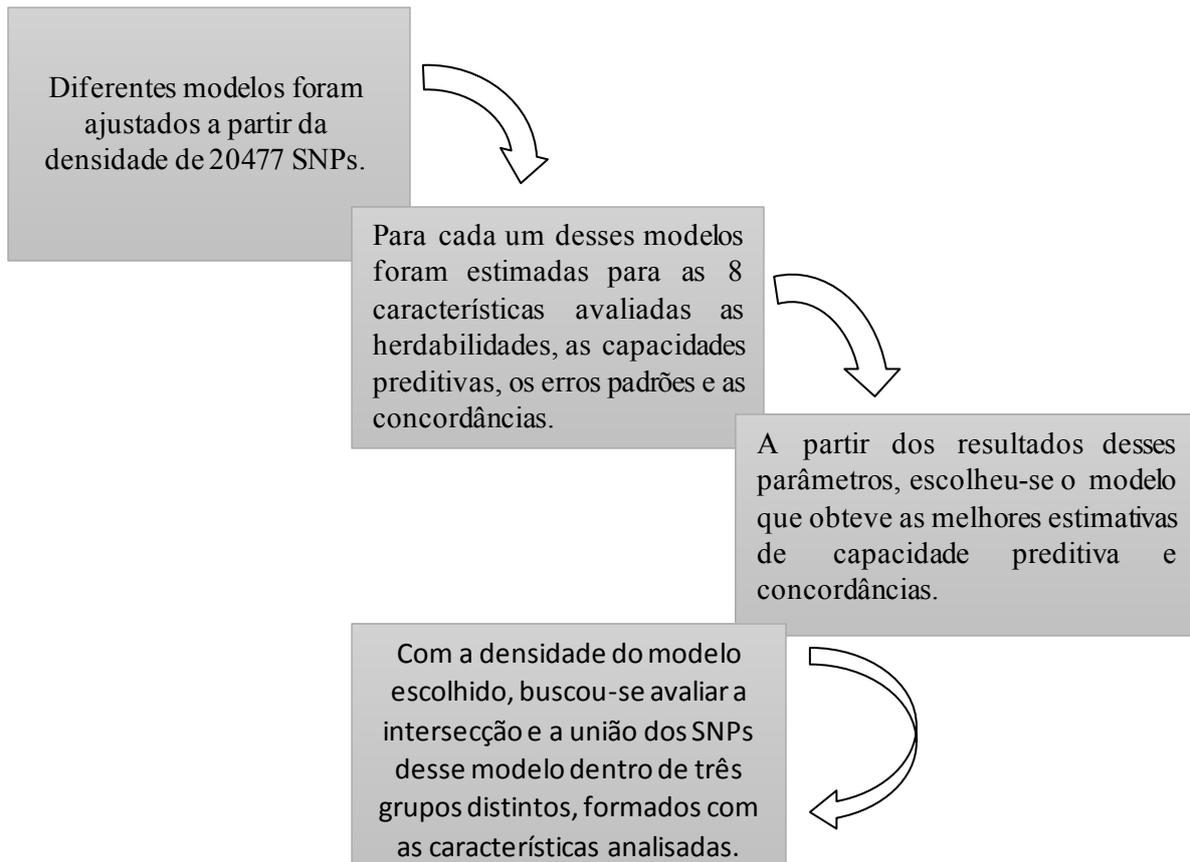


Figura 1: Fluxograma ilustrando as etapas adotadas para a realização das análises.

4. RESULTADOS E DISCUSSÃO

Neste trabalho avaliou-se a viabilidade do uso de painéis de marcadores de baixa densidade na predição do valor genético genômico (\hat{g}) de características economicamente importantes de *C.arábica*. Para isso, utilizou-se o método de seleção genômica G-BLUP, através do qual foram ajustados, para cada característica de interesse, modelos com diferentes densidades de marcadores (2047 até 20477 SNPs), selecionados com base na magnitude de seus efeitos (maiores efeitos em valor absoluto).

Após o ajuste do modelo para as diferentes densidades, a fim de identificar o subconjunto de marcadores que permitisse predizer de forma mais acurada o valor genético genômico (\hat{g}) das características avaliadas, foram estimadas as herdabilidades (h^2) (Tabela 4), as capacidades preditivas ($r_{y\hat{y}}$) e o erro padrão das estimativas de $r_{y\hat{y}}$ (Tabela 5), as concordâncias entre os indivíduos selecionados por meio do valor fenotípico e o do valor genético genômico (\hat{g}) (Tabela 6) e as concordâncias entre os indivíduos selecionados por meio do \hat{g} utilizando todos os 20477 SNPs e aqueles selecionados a partir de modelos ajustados com outras densidades (Tabela 7).

Tabela 4: Herdabilidades estimadas utilizando diferentes densidades de marcadores SNPs.

Característica	Número de SNPs									
	2047	4095	6143	8191	10238	12286	14334	16382	18429	20477
DCo	0,20	0,29	0,41	0,45	0,48	0,49	0,50	0,51	0,53	0,54
CRP	0,20	0,25	0,32	0,37	0,41	0,43	0,45	0,47	0,47	0,48
NNV	0,22	0,35	0,50	0,55	0,57	0,60	0,61	0,61	0,61	0,61
Vig	0,13	0,29	0,38	0,42	0,45	0,46	0,46	0,46	0,45	0,45
TFr	0,18	0,24	0,28	0,30	0,33	0,36	0,38	0,40	0,40	0,41
NFRP	0,33	0,34	0,36	0,38	0,39	0,42	0,44	0,43	0,42	0,40
Cer	0,18	0,23	0,30	0,36	0,41	0,46	0,51	0,53	0,54	0,55
BM	0,12	0,13	0,21	0,24	0,27	0,30	0,33	0,33	0,33	0,32

DCo = Diâmetro da copa; CRP = Comprimento de um ramo plagiotrópico; NNV = Número de nós vegetativos; Vig = Vigor Vegetativo; TFr = Tamanho dos frutos maduros; NFRP = Número de frutos por ramo plagiotrópico; Cer = Incidência de Cercosporiose; BM = Infestação de Bicho-Mineiro.

As estimativas de herdabilidade genômica (h^2) variaram de 0,12 para a característica infestação de bicho-mineiro (BM) para a densidade 2047 SNPs a 0,61 para a característica número de nós vegetativos (NNV) nas densidades de 14334 a 20477 SNPs (Tabela 4). Os valores de h^2 foram de moderados a altos para todas as características avaliadas. Esses resultados mostraram que os genes têm efeito na expressão do fenótipo (COVARRUBIAS-PAZARAN, 2019).

As capacidades preditivas para as diferentes densidades nas oito características avaliadas não obtiveram mudanças consideráveis nas estimativas (Tabela 5), de modo que estatisticamente, considerando-se o erro padrão, elas se mantiveram constantes, mostrando que a predição genômica poderia ser feita considerando o total de SNPs utilizados ou considerando densidades menores. Sousa et al. (2019) em estudos realizados com características de *C. arábica*, obtiveram $r_{y\hat{y}}$ de 0,40 para DCo, 0,32 para CRP, 0,21 para Vig, 0,23 para TFr e 0,31 para Cer, valores estes que estão em conformidade aos que foram encontrados neste estudo.

Tabela 5: Capacidades preditivas médias estimadas e erros padrão das estimativas (entre parênteses) utilizando diferentes densidades de marcadores SNP.

Característica	Número de SNPs									
	2047	4095	6143	8191	10238	12286	14334	16382	18429	20477
DCo	0,36 (0,02)	0,36 (0,02)	0,39 (0,02)	0,39 (0,02)	0,39 (0,02)	0,38 (0,02)	0,37 (0,02)	0,37 (0,02)	0,36 (0,02)	0,36 (0,02)
CRP	0,36 (0,03)	0,35 (0,03)	0,35 (0,03)	0,35 (0,03)	0,35 (0,03)	0,35 (0,03)	0,34 (0,03)	0,34 (0,03)	0,34 (0,03)	0,33 (0,03)
NNV	0,26 (0,07)	0,26 (0,07)	0,29 (0,08)	0,30 (0,08)	0,30 (0,08)	0,30 (0,08)	0,30 (0,07)	0,29 (0,06)	0,28 (0,06)	0,27 (0,05)
Vig	0,30 (0,07)	0,32 (0,08)	0,33 (0,08)	0,33 (0,08)	0,32 (0,08)	0,32 (0,08)	0,30 (0,07)	0,29 (0,07)	0,28 (0,07)	0,27 (0,07)
TFr	0,28 (0,07)	0,27 (0,07)	0,25 (0,07)	0,24 (0,07)	0,24 (0,07)	0,24 (0,07)	0,24 (0,07)	0,23 (0,07)	0,23 (0,07)	0,22 (0,07)

NFRP	0,29 (0,04)	0,26 (0,05)	0,25 (0,05)	0,23 (0,05)	0,21 (0,05)	0,20 (0,04)	0,18 (0,04)	0,16 (0,04)	0,14 (0,03)	0,11 (0,03)
Cer	0,33 (0,08)	0,32 (0,08)	0,31 (0,08)	0,31 (0,08)						
BM	0,12 (0,06)	0,09 (0,06)	0,09 (0,07)	0,09 (0,07)	0,09 (0,07)	0,08 (0,07)	0,08 (0,08)	0,06 (0,08)	0,05 (0,08)	0,03 (0,08)

Com relação aos resultados de concordância entre os indivíduos selecionados, tem-se que o índice Kappa que mede esse parâmetro, variou de 0,31 a 0,62 entre o valor fenotípico e o valor genético genômico predito para as diferentes características avaliadas (Tabela 6). Segundo a classificação apresentada na Tabela 1 proposta por Landis & Koch (1977), esses valores podem ser considerados de razoáveis a muito bons. Além disso, os coeficientes de concordância entre indivíduos selecionados por meio do valor genético genômico predito (\hat{g}) utilizando todos os 20477 marcadores e aqueles selecionados a partir de modelos ajustados com as demais densidades (Tabela 7), variaram de 0,62 a 1, que podem ser considerados de muito bons a excelentes. Os valores de concordâncias, mostraram que ao diminuir a densidade de marcadores, as estimativas para esse parâmetro diminuiriam, porém de modo geral, os indivíduos selecionados continuaram semelhantes.

Diante dos resultados supracitados, percebeu-se que a seleção genômica em *C.arábica* pode ser realizada com uma densidade menor do que 20477 SNPs, uma vez que as capacidades preditivas dos modelos não obtiveram alterações estatisticamente consideráveis e mantiveram bons valores de concordâncias. Dessa forma, a partir da densidade de 2047 marcadores, que corresponde a 10% do total de SNPs disponíveis, procurou-se repetir as análises anteriores, para verificar a possibilidade de se utilizar uma densidade ainda menor sem perda da qualidade da predição.

Tabela 6: Coeficientes de concordância entre indivíduos selecionados por meio do valor fenotípico e do valor genético genômico predito (\hat{g}).

Característica	Cohen's Kappa									
	\hat{g}_{2047}	\hat{g}_{4095}	\hat{g}_{6143}	\hat{g}_{8191}	\hat{g}_{10238}	\hat{g}_{12286}	\hat{g}_{14334}	\hat{g}_{16382}	\hat{g}_{18429}	\hat{g}_{20477}
DCo	0,34	0,34	0,53	0,53	0,56	0,56	0,56	0,56	0,56	0,56
CRP	0,50	0,50	0,50	0,50	0,53	0,53	0,53	0,53	0,53	0,53
NNV	0,34	0,46	0,59	0,59	0,59	0,59	0,59	0,59	0,59	0,59
Vig	0,18	0,34	0,37	0,43	0,47	0,47	0,43	0,43	0,43	0,43
TFr	0,31	0,37	0,40	0,43	0,43	0,43	0,43	0,40	0,43	0,43
NFRP	0,50	0,50	0,56	0,56	0,59	0,59	0,62	0,62	0,62	0,62
Cer	0,31	0,40	0,47	0,53	0,59	0,62	0,62	0,62	0,62	0,62

BM	0,34	0,34	0,50	0,53	0,53	0,59	0,59	0,56	0,53	0,53
----	------	------	------	------	------	------	------	------	------	------

Tabela 7: Coeficientes de concordância entre indivíduos selecionados por meio do valor genético genômico predito (\hat{g}) utilizando todos os 20477 marcadores e aqueles selecionados a partir de modelos ajustados com outras densidades.

		Cohen's Kappa									
		\hat{g}_{2047}	\hat{g}_{4095}	\hat{g}_{6143}	\hat{g}_{8191}	\hat{g}_{10238}	\hat{g}_{12286}	\hat{g}_{14334}	\hat{g}_{16382}	\hat{g}_{18429}	\hat{g}_{20477}
$\hat{g}_{DCo20477}$		0,69	0,72	0,84	0,81	0,87	0,94	0,94	0,97	0,97	1
$\hat{g}_{CRP20477}$		0,87	0,94	0,94	0,97	1	1	1	1	1	1
$\hat{g}_{NNV20477}$		0,65	0,81	0,91	0,94	0,97	0,97	0,97	1	1	1
$\hat{g}_{Vig20477}$		0,65	0,84	0,91	0,94	0,94	0,94	0,97	1	1	1
$\hat{g}_{TFR20477}$		0,62	0,84	0,87	0,91	0,91	0,94	0,94	0,97	1	1
$\hat{g}_{NFRP20477}$		0,78	0,84	0,94	0,94	0,97	0,97	0,97	0,97	0,97	1
$\hat{g}_{Cer20477}$		0,69	0,75	0,81	0,87	0,97	0,97	0,97	0,97	0,97	1
$\hat{g}_{BM20477}$		0,65	0,75	0,84	0,97	0,94	0,91	0,91	0,94	0,97	1

Baseado a relação existente entre as características avaliadas, foram formados três grupos de características, sendo eles compostos pelos traços DCo, CRP, NNV e Vig – Grupo 1, TFr e NFRP – Grupo 2 e Cer e BM – Grupo 3. O Grupo 1 foi formado por características associadas com a morfologia do cafeeiro arábica, o Grupo 2 foi formado por características relacionadas aos frutos de *C.arábica* e o Grupo 3 foi composto por características associadas à doença e praga que afetam o *C.arábica*. A partir dos grupos formados, buscou-se identificar os subconjuntos da intersecção e da união entre os 2047 SNPs de maior magnitude para cada característica dentro de cada grupo. Em seguida, foram avaliadas as capacidades preditivas ($r_{y\hat{y}}$), os erros padrões, as herdabilidades (h^2) e as concordâncias dos modelos baseados nestes subconjuntos.

A intersecção dos 2047 marcadores associados à cada características do Grupo 1 (DCo, CRP, NNV e Vig) resultou em 655 SNPs que corresponde a 32% de 2047 e a 3,2% de 20477. Para esse grupo, os valores estimados de herdabilidade genômica (h^2), variaram de 0,04 para a característica vigor vegetativo (Vig) a 0,13 para a característica de comprimento de um ramo plagiotrópico (CRP) (Tabela 8). Em comparação com a densidade de 2047 SNPs, as herdabilidades para as características do Grupo 1 diminuíram. Esse resultado, pode estar relacionado ao fato de que a variância genotípica seja baixa em comparação com a variância fenotípica para as características avaliadas na densidade de 655 SNPs (COVARRUBIAS-PAZARAN, 2019).

Tabela 8: Herdabilidade genômica (h^2) das características nas diferentes densidades de marcadores dos subconjuntos da intersecção, da união e da densidade de 2047 SNPs.

Grupo	Característica	Modelo Inicial*	Número de SNPs da intersecção				Número de SNPs da união		
		2047	655	890	270	3830	3204	3824	
1	DCo	0,20	0,11	-	-	0,26	-	-	
	CRP	0,20	0,13	-	-	0,22	-	-	
	NNV	0,22	0,06	-	-	0,44	-	-	
	Vig	0,13	0,04	-	-	0,22	-	-	
2	TFr	0,18	-	0,17	-	-	0,18	-	
	NFRP	0,33	-	0,12	-	-	0,36	-	
3	Cer	0,18	-	-	0,10	-	-	0,22	
	BM	0,12	-	-	0,10	-	-	0,10	

A intersecção dos 2047 marcadores do Grupo 2 (TFr e NFRP) resultou em 890 SNPs, que equivale à aproximadamente 43,5% dos 2047 e a 4,3% dos 20477 SNPs. Nesse grupo, a estimativa de herdabilidade para TFr se manteve praticamente constante em comparação com a densidade de 2047 SNPs (Tabela 8). Já para a característica NFRP, a estimativa diminuiu com relação aos 2047 SNPs. No Grupo 3 (Cer e BM), a intersecção dos 2047 SNPs reverteu-se em 270 SNPs que equivale a 13% de 2047 e a pouco mais de 1% de 20477. Neste grupo, Cer e BM obtiveram as mesmas estimativas de herdabilidade, isto é $h^2 = 0,10$. O valor de h^2 diminuiu para a característica Cer e manteve-se praticamente constante para a característica BM em comparação ao valor obtido quando se considerou os 2047 SNPs para cada característica.

As estimativas da capacidade preditiva ($r_{y\hat{y}}$) (Tabela 9) das características do Grupo 1 da intersecção, considerando-se os erros padrões, permaneceram constantes em relação à densidade de 2047 SNPs. Isso mostra que a qualidade da predição não diminuiu com a mudança na densidade de marcadores. Ademais, percebeu-se que as características NNV e Vig que apresentaram os menores valores de herdabilidade (Tabela 8), também não obtiveram mudanças na capacidade de predição do modelo com densidade de 655 SNPs em comparação com a densidade de 2047 SNPs para essas características. Esse resultado corrobora a eficiência da GWS na seleção de traços que apresentam baixa herdabilidade, uma vez que a qualidade de predição não alterou com a diminuição da estimativa desse parâmetro e isso foi demonstrado também por outros autores (LEGARRA et al. 2008; ZHANG et al. 2010; ALKIMIM, 2017).

Com relação ao Grupo 2, a capacidade de predição do modelo contendo 890 SNPs, manteve-se constante em comparação ao modelo contendo 2047 SNPs para a característica TFr

(Tabela 9). O valor encontrado para esta característica foi maior em comparação com o encontrado para a mesma característica em *C.canephora* (ALKIMIM, 2017) que registrou 0,00 de r_{yy} para essa característica. Já para a característica NFRP, houve uma pequena diminuição na estimativa de r_{yy} (Tabela 9). Para o Grupo 3, a característica Cer não registrou, do ponto de vista estatístico, mudança na capacidade preditiva para o modelo contendo 270 SNPs em relação àquele contendo 2047 SNPs, já para a característica BM, houve um ligeiro aumento desse parâmetro. Os valores de r_{yy} encontrados para Cer (0,31) e BM (0,18) também foram registrados por Sousa et al. (2019) em estudos de *C.arábica*.

A união dos 2047 marcadores do Grupo 1 (DCo, CRP, NNV e Vig) resultou em 3830 SNPs (Tabela 8). Para esse grupo, houve aumento da estimativa da herdabilidade para todas as características avaliadas em comparação à densidade de 2047 SNPs. Para o Grupo 2 (TFr e NFRP), a união resultou em 3204 SNPs, nesse grupo não houve mudanças substanciais das estimativas de h^2 . O Grupo 3 composto pelas características Cer e BM, a união dos 2047 SNPs dessas características, resultou em 3824 marcadores, nesse grupo também não houve mudanças consideráveis nas estimativas de herdabilidade.

Tabela 9: Capacidades preditivas médias estimadas do modelo contendo 2047 SNPs e das diferentes densidades da intersecção e da união.

Grupo	Característica	Modelo	Número de SNPs da intersecção				Número de SNPs da união		
		Inicial*	2047	655	890	270	3830	3204	3824
1	DCo	0,36 (0,02)	0,31 (0,02)	-	-	0,35 (0,03)	-	-	
	CRP	0,36 (0,03)	0,32 (0,02)	-	-	0,35 (0,02)	-	-	
	NNV	0,26 (0,07)	0,18 (0,05)	-	-	0,30 (0,08)	-	-	
	Vig	0,30 (0,07)	0,22 (0,07)	-	-	0,30 (0,08)	-	-	
2	TFr	0,28 (0,07)	-	0,28 (0,07)	-	-	0,23 (0,07)	-	
	NFRP	0,29 (0,04)	-	0,18 (0,02)	-	-	0,30 (0,03)	-	
3	Cer	0,33 (0,08)	-	-	0,31 (0,08)	-	-	0,31 (0,08)	
	BM	0,12 (0,06)	-	-	0,18 (0,06)	-	-	0,02 (0,08)	

*Modelo contendo os 10% dos marcadores com melhores estimativas de acurácia e valores de concordância.
(* erro padrão das estimativas.

Com relação às capacidades preditivas, não houve mudanças do ponto de vista estatístico para as características do Grupo 1 da união com relação às estimativas desse parâmetro, mostrando que a qualidade da predição se manteve ao aumentar a densidade de 2047 SNPs para 3830 marcadores. Esse comportamento era esperado, uma vez que, conforme já

apresentado (Tabela 6), foi observado que as densidades entre 2047 SNPs e 20477 SNPs apresentavam estatisticamente mesmos valores de $r_{y\hat{y}}$. Esse comportamento, também foi observado para as características do Grupo 2. Além disso, para esse grupo, a característica NFRP que apresentou a maior estimativa de herdabilidade (0,36), também apresentou a maior estimativa de $r_{y\hat{y}}$ (0,30) (Tabela 9). Em geral, para características com maiores estimativas de herdabilidade é esperado que elas obtenham maiores valores de $r_{y\hat{y}}$, uma vez que esses parâmetros são diretamente proporcionais. Assim como verificado neste estudo, foi observado resposta da capacidade preditiva em relação a herdabilidade em outros estudos de GWS, tais como em *Coffea canephora* (ALKIMIM, 2017), *Anacardium occidentale* (CAVALCANTI et al. 2012) e *Coffea arábica* (SOUSA et al. 2019).

Para o Grupo 3, a estimativa de $r_{y\hat{y}}$ para a densidade de 3824 SNPs permaneceu constante com relação a densidade de 2047 SNPs para a característica Cer (Tabela 9). Em comparação com a característica BM, Cer obteve maior estimativa de h^2 e conseqüentemente, obteve maior estimativa de capacidade preditiva. Outrossim, BM apresentou diminuição de $r_{y\hat{y}}$ para a densidade de 3824 marcadores em comparação com a densidade de 2047 SNPs (Tabela 9).

Com relação à concordância entre os indivíduos selecionados por meio do valor fenotípico e do valor genético genômico predito (\hat{g}) para o Grupo 1 da intersecção, o índice *Kappa* foi razoável para a maioria das características desse grupo (Tabela 10). Percebeu-se que para a característica DCo esse índice foi razoável nas densidades de 2047 e 655 SNPs, enquanto para a característica Vig, esse índice foi ruim para essas duas densidades. Para o Grupo 2, a concordância entre o fenótipo e o GEBV se manteve razoável para a característica TFr ao utilizar a densidade de 890 SNPs correspondente à intersecção das características do Grupo 2. No Grupo 3 da intersecção, as concordâncias permaneceram razoáveis para a densidade de 270 SNPs em comparação com o modelo contendo os 2047 SNPs (Tabela 10). De modo geral, ao utilizar o subconjunto de SNPs correspondentes à intersecção das características dentro de seus respectivos grupos, observou-se que os índices mantiveram suas classificações para a maioria das características avaliadas ao utilizar densidades menores do que 2047 SNPs.

Em relação à concordância entre os indivíduos selecionados por meio do valor fenotípico e do valor genético genômico predito (\hat{g}) para o Grupo 1 da união, o índice *Kappa* variou de razoável a bom para as características desse grupo (Tabela 10). Observou-se um aumento do índice *Kappa* para as características NNV e Vig passando de razoável para bom e

de ruim para razoável, respetivamente. Para o Grupo 2 da união, as concordâncias para as duas características desse grupo permaneceram constantes ao aumentar a densidade de 2047 para 3204 SNPs, o seja, permaneceu razoável para TFr e bom para NFRP (Tabela 10). Esse comportamento também foi observado para as características do Grupo 3 da união, que manteve concordâncias razoáveis para Cer e BM nas densidades de 2047 e 3824 SNPs (Tabela 10).

Tabela 10: Coeficientes de concordância entre indivíduos selecionados por meio do valor fenotípico e do valor genético genômico predito (\hat{g}) para as características dos grupos da união e da intersecção.

Grupo		Cohen's Kappa	Cohen's Kappa			Cohen's Kappa		
		Modelo Inicial*	da intersecção			da união		
		\hat{g}_{2047}	\hat{g}_{655}	\hat{g}_{890}	\hat{g}_{270}	\hat{g}_{3830}	\hat{g}_{3204}	\hat{g}_{3824}
1	DCo	0,34	0,21	-	-	0,31	-	-
	CRP	0,50	0,34	-	-	0,50	-	-
	NNV	0,34	0,18	-	-	0,50	-	-
	Vig	0,18	-0,06	-	-	0,28	-	-
2	TFr	0,31	-	0,28	-	-	0,31	-
	NFRP	0,50	-	0,28	-	-	0,53	-
3	Cer	0,31	-	-	0,21	-	-	0,37
	BM	0,34	-	-	0,31	-	-	0,31

*Modelo contendo os 10% dos marcadores com melhores estimativas de capacidade preditiva e valores de concordância.

Para as concordâncias entre indivíduos selecionados por meio do (\hat{g}) utilizando todos os 20477 SNPs e àqueles selecionados a partir do modelo ajustado com a densidade de 655 marcadores correspondente ao Grupo 1 da intersecção, o índice Kappa diminuiu para as características desse grupo com relação ao modelo contendo os 2047 SNPs, todavia a concordância foi muito boa para CRP, boa para DCo e razoável para NNV (Tabela 11). Para o Grupo 2 da intersecção, o índice Kappa manteve-se 0,62 para a característica TFr para a densidade de 890 SNPs em comparação com a densidade de 2047 marcadores, indicando concordância muito boa. Para o Grupo 3, as concordâncias foram boas para os indivíduos selecionados por meio do GEBV utilizando os 20477 SNPs e àqueles selecionados a partir do modelo com 655 SNPs para as duas características desse grupo.

Tabela 11: Coeficientes de concordância entre indivíduos selecionados por meio do valor genético genômico predito (\hat{g}) utilizando todos os 20477 marcadores e aqueles selecionados a partir do modelo ajustado com as diferentes densidades da intersecção e da união.

Grupo		Cohen's Kappa	Cohen's Kappa			Cohen's Kappa		
		Modelo Inicial*	da intersecção			da união		
		\hat{g}_{2047}	\hat{g}_{655}	\hat{g}_{890}	\hat{g}_{270}	\hat{g}_{3830}	\hat{g}_{3204}	\hat{g}_{3824}
1	$\hat{g}_{DCo20477}$	0,69	0,53	-	-	0,69	-	-
	$\hat{g}_{CRP20477}$	0,87	0,72	-	-	0,91	-	-
	$\hat{g}_{NNV20477}$	0,65	0,34	-	-	0,78	-	-
	$\hat{g}_{Vig20477}$	0,65	0,18	-	-	0,75	-	-
2	$\hat{g}_{TFR20477}$	0,62	-	0,62	-	-	0,75	-
	$\hat{g}_{NFRP20477}$	0,78	-	0,53	-	-	0,87	-
3	$\hat{g}_{Cer20477}$	0,69	-	-	0,53	-	-	0,72
	$\hat{g}_{BM20477}$	0,65	-	-	0,56	-	-	0,65

Para os grupos da união, grande parte das características registraram aumento da estimativa do índice Kappa em comparação ao grupo de 2047. As características DCo, NNV e Vig do Grupo 1 obtiveram concordâncias muito boas, classificação essa registrada também para TFr do Grupo 2, Cer e BM do Grupo 3. O aumento das concordâncias para os grupos da união era esperado, uma vez que quanto maiores forem as densidades, mais elas se aproximam da densidade de 20477 SNPs e conseqüentemente maiores serão as concordâncias entre os indivíduos selecionados por meio do GEBV utilizando esses SNPs e aqueles selecionados a partir do modelo com as diferentes densidades.

Diante do exposto, percebeu-se que a utilização de densidades de marcadores provenientes da intersecção de SNPs dentro de grupos considerando-se a relação entre as características, pode ser uma estratégia interessante para o melhoramento de *C.arábica*, uma vez que ao utilizarmos densidades com menores números de SNPs (655, 890 e 270) as capacidades preditivas dos modelos não apresentaram mudanças consideráveis em comparação com a densidade de 2047 marcadores e conseqüentemente com a densidade de 20477. Além disso, apesar da ligeira diminuição dos índices Kappa para as características desses grupos, os indivíduos continuaram concordando com relação ao fenótipo e com relação ao número total de SNPs. Ademais, observou-se que a utilização da união dos SNPs dentro dos grupos formados, também pode ser uma estratégia viável, já que não houve mudanças substanciais das capacidades preditivas, além das concordâncias entre os indivíduos avaliados terem apresentado notáveis valores. Ao utilizar essas estratégias (intersecção e união de SNPs), o melhorista poderá reduzir tempo e custos de genotipagem no melhoramento de *C.arábica* e isso possibilitará a seleção de cultivares mais produtivas, adaptadas e com qualidade superior de

bebida, pois poderá fazer uso de marcadores com menores densidades, que são tão informativos quanto àqueles correspondentes à densidades maiores.

5. CONCLUSÃO

Os resultados obtidos neste estudo demonstram que o uso de painéis de marcadores de baixa densidade na GWS quando se considera grupos de intersecção e da união entre marcadores importantes para diferentes características de *C. arábica*, pode ser uma ferramenta útil para o melhoramento dessa espécie, uma vez que modelos baseados nestes painéis de menor densidade apresentaram boas estimativas de capacidades preditivas e substanciais valores de concordância em termos de seleção quando comparados à modelos de maior densidade de marcadores.

6. REFERÊNCIAS BIBLIOGRÁFICAS

- ALILLOO, H et al. The feasibility of using low-density marker panels for genotype imputation and genomic prediction of crossbred dairy cattle of East Africa. **Journal of Dairy Science**, v. 101, n.10, p. 9108-9127, out. 2018. DOI: 10.3168/jds.2018-14621.
- ALKIMIM, E. R. Diversidade genética, ganhos com seleção e seleção genômica ampla na espécie *Coffea canéfora*. Tese (Doutorado em Genética e Melhoramento) -Universidade Federal de Viçosa, Viçosa, 2017.
- BANDEIRA e SOUSA, M; GALLI, G.; LYRA, D. H.; GRANATO, I. S. C.; MATIAS, F. I.; ALVES, F. C.; FRITSCHÉ-NETO, R. Increasing accuracy and reducing costs of genomic prediction by marker selection. **Euphytica**, SP, v. 18, p. 215-218, jan. 2019. DOI: <https://doi.org/10.1007/s10681-019-2339-z>.
- BESSADA, S. M. F.; ALVES, R. C.; COSTA, A. S. G.; NUNES, M. A.; OLIVEIRA, B. P. P. *Coffea canephora* silverskin from different geographical origins: A comparative study. **Science of The Total Environment**, v. 645, p. 1021-1028, 2018.
- BHERING, L. L. A tool for biometric and statistical analysis using the R platform. **Crop Breeding and Applied Biotechnology**, v. 17, n. 2, p. 187-190, jun. 2017. DOI: 10.1590/1984-70332017v17n2s29.
- CAETANO, A. R. Marcadores SNP: conceitos básicos, aplicações no manejo e no melhoramento animal e perspectivas para o futuro. **Revista Brasileira de Zootecnia**, Viçosa, v. 38, jul. 2009. DOI: <https://doi.org/10.1590/S1516-35982009001300008>.
- CAIXETA, E. T.; PESTANA, K. N.; PESTANA, R. K. N. Melhoramento do cafeeiro: ênfase na aplicação dos marcadores moleculares. In: GARCIA, G.O.; REIS, E. F.; LIMA, J. S. S.; XAVIER, A. C.; RODRIGUES, W. N.; **Tópicos especiais em produção vegetal V**. Alegre, ES: CAUFES, 2015. p. 154-179.
- CARVALHO, C. H. S. **Cultivares de Café**: origem, características e recomendações. Brasília, DF: Embrapa Café, 2008. Disponível em: <https://livimagens.sct.embrapa.br/amostras/00082810.pdf>. Acesso em: 16 dez. 2020.
- CAVALCANTI, J. J. V.; RESENDE, M. D. V. de.; SANTOS, F. H. C. dos.; PINHEIRO, C. R. Predição simultânea dos efeitos de marcadores moleculares e seleção genômica ampla em cajueiro. **Revista Brasileira de Fruticultura**, v. 34, n. 3, p. 840-846, set. 2012. DOI: <https://doi.org/10.1590/S0100-29452012000300025>.
- CECAFE Conselho dos Exportadores de Café do Brasil. Disponível em: http://www.sapc.embrapa.br/arquivos/consorcio/informe_estatistico/CECAFE_Relatorio_Mensal_SETEMBRO_2020.pdf. Acesso em: 11 nov. 2020.
- CECAFE Conselho dos Exportadores de Café do Brasil. Disponível em: http://www.sapc.embrapa.br/arquivos/consorcio/informe_estatistico/CECAFE_Relatorio_Mensal_OUTUBRO_2019.pdf. Acesso em: 11 nov. 2020.
- CLARK, S. A. WERF, J. van der. Genomic Best Linear Unbiased Prediction (gBLUP) for the Estimation of Genomic Breeding Values. **Methods in molecular biology**, v. 1019, p. 321-330, 2013. DOI: 10.1007/978-1-62703-447-0_13.

- COHEN, J. A coefficient of agreement for nominal scales. **Educational and Psychological Measurement**, v. 20, p. 37-46, abr. 1960. DOI: <https://doi.org/10.1177/001316446002000104>.
- COVARRUBIAS-PAZARAN, G. E. **Heritability: meaning and computation**. Breeding Optimization Lead, CGIAR Excellence in Breeding Platform (EiB), dez. 2019. Disponível em: https://excellenceinbreeding.org/sites/default/files/manual/EiB-M2_Heritability_18-02-20.pdf.
- DESTA, Z. A.; ORTIZ, R. Genomic selection: genome-wide prediction in plant improvement. **Trends in Plant Science**, v.19, p. 592-601, set. 2014. DOI: <https://doi.org/10.1016/j.tplants.2014.05.006>.
- ENDELMAN, J. B. Ridge regression and other kernels for genomic selection with R package rrBLUP. **Plant Genome**, USA, v.4, p. 250-255, nov. 2011. DOI: <https://doi.org/10.3835/plantgenome2011.08.0024>.
- FERNANDO, R. L. *et al.* Genomic selection. **Acta Agriculturae Scandinavica, Section A - Animal Science**, v. 57, p. 192-195, abr. 2008. DOI: <https://doi.org/10.1080/09064700801959395>.
- FIESP. Disponível em: <https://www.fiesp.com.br/indices-pesquisas-e-publicacoes/balanco-comercial/>. Acesso em: 29 jun. 2021.
- FILHO, H. P. M.; BORDIGNON, R. Melhoramento do café arábica ao longo dos anos. *In*: CARVALHO, C. H. S. **Cultivares de café**. Brasília: EMBRAPA, 2007. p. 65-82.
- FOODCHAIN. Disponível em: <https://www.bing.com/search?q=principais+produtos+agr%C3%ADcolas+de+expota%C3%A7%C3%A3o+do+brasil&qs=n&form=QBRE&sp=1&pq=principais+produtos+agr%C3%ADcolas+de+expota%C3%A7%C3%A3o+do+brasil&sc=0-52&sk=&cvid=091A615CF54D47A3A1D3A27DC77AFEB2#>. Acesso em: 08 nov. 2020.
- FREITAS, Z. M. T. S.; OLIVEIRA, F. J.; CARVALHO, S. P.; SANTOS, V. F.; SANTOS, J. P. Avaliação de caracteres quantitativos relacionados com o crescimento vegetativo entre cultivares de café arábica de porte baixo. **Bragantia**, v.66, p.267-275, 2007.
- GODDARD, M. E.; HAYES, B. J. Genomic selection. **Journal Breeding and Genetics**, v. 124, n.6, p. 323-330, out. 2007. DOI: <https://doi.org/10.1111/j.1439-0388.2007.00702.x>.
- GOIS, I. B *et al.* Genome wide selection in Citrus breeding. **Genetics and Molecular Research**, v. 15, n. 4, out. 2016. DOI: [10.4238/gmr15048863](https://doi.org/10.4238/gmr15048863).
- GRANATO, I. S. C *et al.* Seleção de marcadores para métodos RR-BLUP e BLASSO na seleção genômica ampla. *In*: Congresso Brasileiro de Melhoramento de Plantas, 7., 2013, Uberlândia. **Anais ... Uberlândia**: Embrapa, 2013. p. 285-288.
- GRATTAPAGLIA, D.; RESENDE, M. D. V. de. Genomic selection in forest tree breeding. **Tree Genetics & Genomes**, v. 7, p. 241-255, out. 2010. DOI: [10.1007/s11295-010-0328-4](https://doi.org/10.1007/s11295-010-0328-4).
- HABIER, D.; FERNANDO, R. L.; DEKKERS, J. C. M. The impact of genetic relationship information on genome-assisted breeding values. **Genetics**, v. 177, n. 4, p. 2389-2397, dez. 2007. DOI: <https://doi.org/10.1534/genetics.107.081190>.
- HENDERSON, C. R. Sire evaluation and genetic trends. **Journal of Animal Science**, v. 1973, p. 10-41, jan. 1973. DOI: <https://doi.org/10.1093/ansci/1973.Symposium.10>.

- HICKEY, J. M.; CHIURUGWI, T.; MACKAY, I.; POWELL, W. Genomic prediction unifies animal and plant breeding programs to form platforms for biological discovery. **Nature Genetics**, v.49, n.9, p.1297-1303, ago. 2017. DOI: 10.1038/ng.3920.
- HIDALGO, A. M et al. Accuracy of predicted genomic breeding values in purebred and crossbred pigs. **G3: GENES, GENOMES, GENETICS**, v.5, n.8, p. 1575-1583, ago. 2015. DOI: <https://doi.org/10.1534/g3.115.018119>.
- HOFFSTETTER, A.; CABRERA, A.; HUANG, M.; SNELLER, C. Optimizing training population data and validation of genomic selection for economic traits in soft winter wheat. **G3: GENES, GENOMES, GENETICS**, Sociedade de Genética da América, v. 6, n. 9, p. 2919-2928, set. 2016. DOI: <https://doi.org/10.1534/g3.116.032532>.
- JAVAREMI, A. N.; SMITH, C. GIBSON, J. P. Effect of total allelic relationship on accuracy of evaluation and response to selection. **Journal of Animal Science**, v. 75, n.7, p. 1738-1745, jul. 1997. DOI: 10.2527/1997.7571738x.
- KOHAVI, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. **IJCAI'95**, v.2, p.1137-1143, ago. 1995.
- KRIARIDOU, C.; TSAIRIDOU, S.; HOUSTON, R. D.; ROBLEDO, D. Genomic prediction using low density marker panels in aquaculture: performance across species, traits, and genotyping platforms. **Frontiers in Genetics**, Espanha, v. 11, p. 124, fev. 2020. DOI: <https://doi.org/10.3389/fgene.2020.00124>.
- LANDIS, J.; KOCH, G. The measurement of observer agreement for categorical data. **Biometrics**, v. 33, n. 1, p.159-174, mar.1977.
- LEGARRA, A.; ROBERT-GRANIÉ, C.; MANFREDI, E.; ELSÉN, J-M. Performance of Genomic Selection in Mice. **Genetics**, v.180, n.1, p.611-618, set. 2008. DOI: <https://doi.org/10.1534/genetics.108.088575>.
- MAIA, T et al. The *Hemileia vastatrix* effector HvEC-016 suppresses bacterial blight symptoms in coffee genotypes with the SH1 rust resistance gene. **New Phytologist**, v. 213, p. 1315-1329, dez. 2016. DOI: <https://dx.doi.org/10.1111%2Fnph.14334>.
- MAPA. Disponível em: <https://www.gov.br/agricultura/pt-br/assuntos/politica-agricola/cafe/cafeicultura-brasileira>. Acesso em: 19 jan. 2021.
- MEUWISSEN, T.H.E.; HAYES, B.J.; GODDARD, M.E. Prediction of total genetic value using genome wide dense marker maps. **Genetics**, v.157, p. 1819–1829, 2001.
- MOSS, S. R. **Competitividade da produção do café arábica em Minas Gerais e São Paulo**. 2006. 91p. Tese (Mestrado em Economia Aplicada) – Universidade Federal de Viçosa, Viçosa, 2006.
- OLIVEIRA, A. C. B et al. Prediction of genetic gains from selection in Arabica coffee progênies. **Crop Breeding and Applied Biotechnology**, Viçosa, v. 11, n. 2, p. 106-113, jun. 2011. DOI: <https://doi.org/10.1590/S1984-70332011000200002>.
- PEREIRA e OLIVEIRA. Cultivares. In: SAKIYAMA, N. S.; MARTINEZ, H. E. P.; TOMAZ, M. A.; BORÉM, A. **Café arábica: do plantio à colheita**. Viçosa: UFV, 2015. p. 316.

PETEK, M. R.; SERA, T.; FONSECA, I. C. B. de. Exigências climáticas para o desenvolvimento e maturação dos frutos de cultivares de Coffea arábica. *Bragantia*, Campinas, v. 68, n.1, p. 169-181, 2009. DOI: <http://dx.doi.org/10.1590/S0006-87052009000100018>.

PORTAL DO CAFÉ DE MINAS. Disponível em: <http://portaldocafedeminas.emater.mg.gov.br/>. Acesso em: 16 dez. 2020.

RESENDE, M. D. V. de. Software Selegen-REML/BLUP: a useful tool for plant breeding. *Crop Breeding and Applied Biotechnology*, v. 16, n. 4, p. 330-339, dez. 2016. DOI: <https://doi.org/10.1590/1984-70332016v16n4a49>.

R CORE TEAM. A language and environment for statistical computing. **R**, 2019.

RESENDE, M. D. V. de.; BARBOSA, M. H. P.; REZENDE, G. D. S. P.; AGUIAR, A. M.; DIAS, L. A. dos S.; STURION, J. A. Métodos e estratégias de melhoramento de espécies perenes: estado da arte e perspectivas. *In: CONGRESSO BRASILEIRO DE MELHORAMENTO DE PLANTAS*, 3., 2005, Gramado. **Anais**. Passo Fundo: Embrapa Trigo; [S.l.]: Sociedade Brasileira de Melhoramento de Plantas, 2005.

RESENDE, M. D. V. de.; JÚNIOR, E. F.; MORAES, M. L. T.; FAZUOLI, L. C. Estimativas de parâmetros genéticos e predição de valores genotípicos no melhoramento do cafeeiro pelo procedimento REML/BLUP. *Bragantia*, Campinas, SP, v. 60, n. 3, p. 185-193, nov. 2001. DOI: <http://dx.doi.org/10.1590/S0006-87052001000300005>.

RESENDE, M. D. V. de.; LOPES, P. S.; SILVA, R. L.; PIRES, I. E. Seleção genômica ampla (GWS) e maximização da eficiência do melhoramento genético. *Pesquisa Florestal Brasileira*, Colombo, n. 56, p. 63-78, jan./jun. 2008.

RESENDE, M. D. V. de.; SILVA, F. F.; AZEVEDO, C. F. **Estatística matemática, biométrica e computacional: Modelos Mistos, Multivariados, Categóricos e Generalizados (REML/BLUP), Inferência Bayesiana, Regressão Aleatória, Seleção Genômica, QTL-GWAS, Estatística Espacial e Temporal, Competição, Sobrevivência**. Viçosa: Suprema, 881p. 2014.

RESENDE, M. D. V.; DUARTE, J. B. Precisão e controle de qualidade em experimentos de avaliação de cultivares. *Pesquisa Agropecuária Tropical*, v. 37, n. 3, p. 182–194, set. 2007.

RESENDE, M. D.V. de.; SILVA, F.F.; LOPES, P.S.; AZEVEDO, C.F. **Seleção Genômica Ampla (GWS) via Modelos Mistos (REML/BLUP), Inferência Bayesiana (MCMC), Regressão Aleatória Multivariada e Estatística Espacial**. Viçosa: Universidade Federal de Viçosa/Departamento de Estatística. 2012.

SAKIYAMA, N. S. O café arábica. *In: SAKIYAMA, N. S.; MARTINEZ, H. E. P.; TOMAZ, M. A.; BORÉM, A. Café arábica: do plantio à colheita*. Viçosa: UFV, 2015. p. 316.

SERA, T. Coffee genetic breeding at IAPAR. *Crop Breeding and Applied Biotechnology*, v.1, n.2, p.179-190, jun. 2001. DOI: [10.13082/1984-7033.v01n02a08](https://doi.org/10.13082/1984-7033.v01n02a08).

SOUSA, T. V et al. Early Selection Enable by the implementation of Genomic Selection in Coffea arábica Breeding. *Frontiers in plant Science*, v. 8, n.9, jan. 2019. DOI: [10.3389/fpls.2018.01934](https://doi.org/10.3389/fpls.2018.01934).

SPINDEL, J et al. Genomic selection and association mapping in rice (*Oryza sativa*): effect of trait genetic architecture, training population composition, marker number and statistical model

on accuracy of rice genomic selection in elite, tropical rice breeding lines. **Plos Genetics**. Estados Unidos, v.11, n. 2, fev. 2015. DOI: <https://doi.org/10.1371/journal.pgen.1004982>.

SUMÁRIO EXECUTIVO DO CAFÉ. Disponível em: http://www.sapc.embrapa.br/arquivos/consorcio/informe_estatistico/Sumario_Cafe_Novembro_2020.pdf. Acesso em 27 jan. 2021.

TALHINHAS, P et al. The coffee leaf rust pathogen *Hemileia vastatrix*: one and a Half centuries around the tropics. **Molecular Plant Pathology**, v. 18, n. 8, p. 1039-1051, jan. 2017. DOI:10.1111/mpp.12512.

TEIXEIRA, F. R. F. **ANÁLISE DE FATORES APLICADA NA SELEÇÃO GENÔMICA EM SUÍNOS**. 2015. 63p. Dissertação (Mestrado em Estatística Aplicada e Biometria) – Universidade Federal de Viçosa, Viçosa, 2015.

VAN DER VOSSSEN, H.; BERTRAND, B.; CHARRIER, A. Next generation variety development for sustainable production of arabica coffee (*Coffea arabica* L.): a review. **Euphytica**, v. 204, p. 243-256, fev. 2015. DOI: <https://doi.org/10.1007/s10681-015-1398-z>.

VANRADEN, P. M. Efficient methods to compute genomic predictions. **Jornal of Dairy Science**, v.91, p. 4414-4423, nov. 2008. DOI: <https://doi.org/10.3168/jds.2007-0980>.

VAZQUEZ, A. I et al. Predictive ability of subsets of single nucleotide polymorphisms with and without parent average in US Holsteins. **Jornal of Dairy Science**, v.93, p. 5942-5949, nov. 2010. DOI: <https://dx.doi.org/10.3168%2Fjds.2010-3335>.

VIDAL, R. O et al. A high-throughput data mining of single nucleotide polymorphisms in coffee species expressed sequence tags suggests differential homeologous gene expression in the allotetraploid coffee arabica. **Plant Physiology**, v.154, n.3, p. 1053-1066, nov. 2010. DOI: 10.1104/pp.110.162438.

VILLANUEVA, B.; PONG-WONG, R.; FERNÁNDEZ, J.; TORO, M. A. Benefits from marker-assisted selection under an additive polygenic genetic model. **Journal of Animal Science**, v.83, n.8, p. 1747-1752, ago. 2005. DOI: <https://doi.org/10.2527/2005.8381747x>.

WELLMANN, R et al. Genomic selection using low density marker panels with application to a sire line in pigs. **Genetics Selection Evolution**, v.45, n.28, jul. 2013. DOI: <https://doi.org/10.1186/1297-9686-45-28>.

WHITTAKER, J. C.; THOMPSON, R.; DENHAM, M. C. Marker assisted selection using ridge regression. **Genetics Research**, Cambridge, v.75, p. 249-252, abr. 2000. DOI: <https://doi.org/10.1017/S0016672399004462>.

ZAMBOLIM, L. Current status and management of coffee leaf rust in Brazil. **Tropical Plant Pathology**, v. 41, p. 1-8, jan. 2016. DOI: <https://doi.org/10.1007/s40858-016-0065-9>.

ZHANG, H.; YIN, L.; WANG, M.; YUAN, X.; LIU, X. Factors affecting the accuracy of genomic selection for agricultural economic traits in maize, cattle, and pig populations. **Frontiers in Genetics**, Espanha, v.10, p. 189, mar. 2019. DOI: 10.3389/fgene.2019.00189.

ZHANG, Z et al. Accuracy of whole-genome prediction using a genetic architecture-enhanced variance-covariance matrix. **G3: GENES, GENOMES, GENETICS**, Sociedade de Genética da América, v. 5, n. 4, p. 614-627, fev. 2015. DOI: <https://doi.org/10.1534/g3.114.016261>.

ZHANG, Z. *et al.* Mixed linear model approach adapted for genome wide association studies. **Nature Genetics**, v. 42, n. 4, p. 355-360, mar. 2010. DOI: 10.1038/ng.546.