

**SELEÇÃO RECORRENTE GENÔMICA NO MELHORAMENTO
DE MILHO-PIPOCA**

ISMAEL ALBINO SCHWANTES

**UNIVERSIDADE ESTADUAL DO NORTE FLUMINENSE
DARCY RIBEIRO – UENF**

**CAMPOS DOS GOYTACAZES – RJ
AGOSTO – 2019**

SELEÇÃO RECORRENTE GENÔMICA NO MELHORAMENTO
DE MILHO-PIPOCA

ISMAEL ALBINO SCHWANTES

“Tese apresentada ao Centro de Ciências e
Tecnologias Agropecuárias da Universidade
Estadual do Norte Fluminense Darcy Ribeiro,
como parte das exigências para obtenção do
título de Doutor em Genética e Melhoramento
de Plantas”

Orientador: Prof. Antônio Teixeira do Amaral Junior

CAMPOS DOS GOYTACAZES – RJ
AGOSTO-2019

FICHA CATALOGRÁFICA

UENF - Bibliotecas

Elaborada com os dados fornecidos pelo autor.

S398

Schwantes, Ismael Albino.

Seleção recorrente genômica no melhoramento de milho-pipoca / Ismael Albino Schwantes. - Campos dos Goytacazes, RJ, 2019.

74 f. : il.

Bibliografia: 45 - 59.

Tese (Doutorado em Genética e Melhoramento de Plantas) - Universidade Estadual do Norte Fluminense Darcy Ribeiro, Centro de Ciências e Tecnologias Agropecuárias, 2019.

Orientador: Antonio Teixeira do Amaral Junior.

1. seleção genômica. 2. seleção recorrente. 3. milho-pipoca. 4. ganho genético. 5. marcadores SNPs. I. Universidade Estadual do Norte Fluminense Darcy Ribeiro. II. Título.

CDD - 631.5233

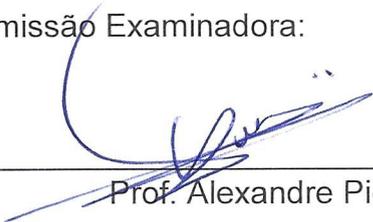
SELEÇÃO RECORRENTE GENÔMICA NO MELHORAMENTO
DE MILHO-PIPOCA

ISMAEL ALBINO SCHWANTES

“Tese apresentada ao Centro de Ciências e
Tecnologias Agropecuárias da Universidade
Estadual do Norte Fluminense Darcy Ribeiro,
como parte das exigências para obtenção do
título de Doutor em Genética e Melhoramento de
Plantas”

Aprovada em 21 de agosto de 2019.

Comissão Examinadora:



Prof. Alexandre Pio Viana (D.Sc., Produção Vegetal) - UENF



Prof. Messias Gonzaga Pereira (Ph.D., Plant Breeding) - UENF



Dr. Janeo Eustáquio de Almeida Filho (D.Sc., Genética e Melhoramento) - Bayer
Coxilha-RS



Prof. Antônio Teixeira do Amaral Junior (D.Sc., Genética e Melhoramento) - UENF
(Orientador)

DEDICATÓRIA

Aos meus pais Clautelio e Marlene, aos meus irmãos Diogo e Luan, a minha esposa Andreia e minha filha Yasmin.

AGRADECIMENTOS

Ao nosso bom DEUS, pelo dom da vida e pela presença marcante em minha vida.

Especialmente tenho o dever e a satisfação de agradecer: a toda minha família, alicerce de todo o meu ser, em especial aos meus pais, Clautelio e Marlene, à minha avó Irma, e aos meus irmãos Diogo e Luan.

À minha eterna namorada Andréia, pelo amor, amizade e companheirismo, sempre fundamentais para a realização dos sonhos planejados em conjunto. A minha filha Yasmin, por me fazer conhecer o maior amor do mundo. Aos amigos para todo sempre Renan, Pelé, Laércio, Ezequiel, Jorge e João, pela amizade e “parceria” ímpares.

À Universidade Estadual do Norte Fluminense Darcy Ribeiro e ao Programa de Pós-Graduação em Genética e Melhoramento de Plantas, por possibilitar minha formação acadêmica em nível de Doutor em Genética e Melhoramento de Plantas.

Ao meu orientador Prof. Dr. Antônio Teixeira do Amaral Junior pela orientação e ensinamentos concedidos durante a execução do projeto, por acreditar em mim e muito contribuir para meu crescimento profissional e pela oportunidade de realizar este trabalho. Agradeço imensamente pela amizade e valorosos conselhos no âmbito pessoal.

Ao Programa de Pós-Graduação em Genética e Melhoramento de Plantas pelo convívio e aprendizado. Agradeço aos Professores que contribuíram com a formação acadêmica, sobretudo a Antônio Teixeira do Amaral Junior, Alexandre Pio

Viana, Messias Gonzaga Pereira, Gonçalo Apolinário de Souza Filho e Marcelo Vivas.

Aos colegas do Grupo de Melhoramento e amigos que participaram deste trabalho e pelos momentos de convívio. Especialmente para Janeo Eustáquio de Almeida, Pedro Diniz, Guilherme Ferreira Pena, Fernando Higinio, Railan Kurosawa, Gabrielle Mafra, Juliana Santos, Yure Pequeno de Souza, Shahid Khan, Ismael Schegoscheski Gerhardt, Adriano dos Santos, Gabriel Moreno, José Arantes, Derivaldo da Cruz, Fernando Ferreira, Jhean Torres Leite, Fábio Tomas e Samuel Kamphorst.

Ao secretário do Programa de Pós-Graduação em Genética e Melhoramento de Plantas, José Daniel Valle de Almeida, agradeço por toda atenção que dispensou durante o meu Doutorado.

Aos técnicos de campos da UENF, Geraldo Francisco de Carvalho, Valdinei Cruz de Azevedo e Luzimara Lima Pereira por toda a colaboração e conselhos na condução dos experimentos.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior pela concessão das bolsas.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

Por final, agradeço a todos que fizeram parte desta caminhada, contribuindo, física e espiritualmente na concretização desta importante etapa em minha vida pessoal e profissional.

Muito obrigado!

SUMÁRIO

LISTA DE FIGURAS	vii
LISTA DE TABELAS	viii
RESUMO	ix
ABSTRACT	xii
1. INTRODUÇÃO	1
2. OBJETIVOS	4
2.1. OBJETIVO GERAL.....	4
2.2. OBJETIVOS ESPECÍFICOS.....	4
3. CAPÍTULOS	5
3.1. SELEÇÃO GENÔMICA PROPORCIONANDO MAIOR GANHO GENÉTICO ANUAL NO MELHORAMENTO POPULACIONAL DE MILHO-PIPOCA	5
3.1.1. INTRODUÇÃO	5
3.1.2. REVISÃO	6
3.1.2.1. MELHORAMENTO DO MILHO-PIPOCA NA UENF	6
3.1.2.2. SELEÇÃO GENÔMICA	8
3.1.3. MATERIAL E MÉTODOS.....	10
3.1.3.1. POPULAÇÃO DE ESTUDO, GENOTIPAGEM E FENOTIPAGEM..	10
3.1.3.2. MÉTODO ESTATÍSTICO APLICADO NA SELEÇÃO GENÔMICA.....	12
3.1.3.3. ESTIMATIVAS DA ACURÁCIA SELETIVA E GANHO DE SELEÇÃO.....	13

3.1.4. RESULTADOS E DISCUSSÃO	14
3.1.4.1 POPULAÇÃO DE ESTUDO E PARÂMETROS GENÉTICOS.....	14
3.1.4.2. ACURÁCIA NA SELEÇÃO GENÔMICA.....	18
3.1.4.3. CORRELAÇÕES GENÉTICA E AMBIENTAL E GANHO GENÉTICO.....	19
3.1.5. CONCLUSÕES.....	24
3.2. ANÁLISE COMPARATIVA DE ALGORITMOS BAYESIANOS E FREQUENTISTAS NA EFICIÊNCIA DA SELEÇÃO RECORRENTE GENÔMICA EM MILHO-PIPOCA E OTIMIZAÇÃO DE DENSIDADES DE MARCADORES SNPs.....	25
3.2.1. INTRODUÇÃO.....	25
3.2.2. REVISÃO.....	27
3.2.2.1. MÉTODOS ESTATÍSTICOS APLICADOS NA GS	27
3.2.2.2. USO E CONTROLE DE QUALIDADE DE SNPS	29
3.2.2.3. DESEQUILÍBRIO DE LIGAÇÃO E ESTRUTURA POPULACIONAL	30
3.2.3. MATERIAL E MÉTODOS.....	31
3.2.3.1. POPULAÇÃO DE ESTUDO.....	31
3.2.3.2. GENOTIPAGEM.....	31
3.2.3.3. FENOTIPAGEM.....	32
3.2.3.4. MÉTODOS ESTATÍSTICOS E VALIDAÇÃO CRUZADA.....	33
3.2.3.5. 'CORRELATION NETWORK' E TEMPOS DE PROCESSAMENTO COMPUTACIONAL.....	37
3.2.3.6. DIFERENTES DENSIDADES DE MARCADORES SNPS ..	37
3.2.4. RESULTADOS E DISCUSSÃO	38
3.2.4.1. COMPARAÇÃO DE MÉTODOS.....	38
3.2.4.2. DIFERENTES DENSIDADES DE SNPs.....	42
3.2.5. CONCLUSÕES.....	44
REFERÊNCIAS BIBLIOGRÁFICAS.....	45

LISTA DE FIGURAS

Figura 1 - Aspectos genéticos da população de estudo.....	10
Figura 2 - Gráficos das correlações genéticas e residuais obtidas.....	21
Figura 3 - Ganho genético anual para as diferentes estratégias utilizadas.....	23
Figura 4 - Estimativas das acurácias preditivas das médias obtidas para os métodos Bayes A, Bayes B, Bayes C, Bayes Lasso, GBLUP e RKHS.....	39
Figura 5 - ' <i>Correlation networks</i> ' obtidas nos métodos GBLUP, Bayes A, Bayes B, Bayes C, Bayes Lasso e RKHS.....	41
Figura 6 - Gráficos que expressam as diferentes acurácias.....	43

LISTA DE TABELAS

Tabela 1 - Ciclos de seleção recorrente na população UNB-2U.....	8
Tabela 2 - Estimativas da acurácia seletiva considerando apenas o fenótipo (FEN, considerando fenótipo e marcas (FEN + GEN simultaneamente e considerando apenas marcas (GEN; componentes de variância aditiva (V_a), da interação dos efeitos aditivos x locais (V_{al}), e herdabilidade média para as características avaliadas.....	18
Tabela 3 - Principais métodos estatísticos para predição genômica.....	28
Tabela 4 - Estimativa de tempo gasto para obtenção de resultados pelos métodos Bayes A, Bayes B e GBLUP.....	42

RESUMO

SCHWANTES, Ismael Albino; D.Sc.; Universidade Estadual do Norte Fluminense Darcy Ribeiro. Agosto, 2019. Seleção Recorrente Genômica no Melhoramento de Milho-Pipoca. Orientador: Antônio Teixeira do Amaral Junior. Conselheiros: Alexandre Pio Viana, Messias Gonzaga Pereira e Janeo Eustáquio de Almeida Filho.

A Seleção Recorrente (SR) é uma eficiente estratégia de melhoramento populacional, em que a população melhorada pode ser recomendada como cultivar e, ou ser base para obtenção de linhagens elite. Porém, este procedimento é trabalhoso e demanda muito tempo, pois a cada ciclo de seleção geralmente são realizadas três etapas: obtenção das progênies, avaliação e recombinação das famílias superiores. Com o intuito de aumentar a eficiência da seleção, foi proposta a utilização da Seleção Recorrente Genômica (SRG), que possibilita realizar as fases de avaliação e recombinação simultaneamente, além de proporcionar uma forma de seleção precoce direta. Na seleção genômica são utilizados painéis densos de marcadores SNPs e vários métodos podem ser empregados para obter a predição da acurácia seletiva. O estudo foi realizado com a população UENF-14, obtida após o oitavo ciclo de Seleção Recorrente (SR) de Milho-Pipoca. Após a recombinação das progênies selecionadas, foram semeadas as sementes da população resultante, iniciando o nono ciclo de SR, o qual foi utilizado neste trabalho. Foram amostrados 200 indivíduos para a genotipagem sendo que desses, 98 indivíduos foram autofecundados, para avaliação fenotípica de suas progênies

S₁. Os ensaios de campo foram realizados em dois locais, sob delineamento de blocos incompletos com três repetições. Os genitores dessas progênes foram genotipados com um painel de 21.442 SNPs, e utilizando o software Plink, foram realizados três filtros, na seqüência: a) retirada de indivíduos com > 10 % de dados perdidos; b) descarte de SNP com > 5 % de dados perdidos; e c) descarte de SNP com menor frequência alélica-MAF < 5%, resultando em 10.507 SNPs. O primeiro capítulo teve como objetivo obter as acurácias seletivas e ganhos genéticos para diferentes estratégias de seleção: FEN= estimativas obtidas somente com as informações fenotípicas de 98 indivíduos; FEN + GEN= estimativas obtidas com as informações fenotípicas e genotípicas de 98 indivíduos; e GEN= estimativas obtidas somente com a genotipagem via marcadores SNPs. Foram avaliados: peso de 100 grãos (P100), altura de espiga (AE), rendimento de grãos (RG), capacidade de expansão (CE), altura de planta (AP) e volume de pipoca (VP). A partir dos resultados obtidos com as predições utilizando diferentes intensidades de seleção para a estratégia GEN foi possível observar que o ganho genético anual médio para as diferentes características foi superior às estratégias FEN e GEN+FEN em 29,1 % e 25,2 % para 98 candidatos à seleção; em 148,3 % e 140,9 % para 500; e em 187,9 % e 179,4 % para 1.000 candidatos a seleção, respectivamente. Isto posto, o uso da seleção recorrente genômica pode representar um elevado ganho genético, desde que: a) a fenotipagem seja realizada de modo preciso; b) seja explorada a intensidade de seleção com a genotipagem de vários indivíduos, aumentando o número de candidatos à seleção, e c) utilize-se a seleção genômica para seleção precoce na SR. No segundo capítulo, o objetivo foi obter a acurácia seletiva utilizando os seguintes métodos: GBLUP, Bayes A, Bayes B, Bayes Cπ, Bayes Lasso e RKHS, e após isso, definir o melhor método avaliando diferentes densidades de marcadores SNPs. O rendimento de grãos (RG), volume de pipoca (VP) e a capacidade de expansão (CE) foram avaliados. O método GBLUP apresentou resultados semelhantes aos demais e foi escolhido para a próxima etapa, devido a sua simplicidade de execução, menor tempo demandado na análise e maior precisão numérica para a característica CE. A partir deste ponto, vários cenários com redução do painel de SNPs foram testados usando o software R. Nos diferentes cenários de densidade de SNPs, os resultados demonstraram que o uso de ~4.800 SNPs apresentou precisão seletiva semelhante ao uso de 10.507 SNPs.

Conclui-se, então, que uma considerável redução de custos pode ser obtida usando um painel reduzido de SNPs e que o método GBLUP deve ser utilizado.

Palavras-chave: seleção genômica; acurácia; marcadores SNPs; milho-pipoca.

ABSTRACT

SCHWANTES, Ismael Albino; D.Sc.; Universidade Estadual do Norte Fluminense Darcy Ribeiro. Agosto, 2019. Recurrent Genomic Selection in Popcorn Breeding Orientador: Antônio Teixeira do Amaral Junior. Conselheiros: Alexandre Pio Viana, Messias Gonzaga Pereira e Janeo Eustáquio de Almeida Filho.

Recurrent Selection (SR) is an efficient population breeding strategy in which the improved population can be recommended as a cultivar and/or a basis for obtaining elite lines. However, this procedure is laborious and time-consuming, since each selection cycle usually involves three stages: obtaining progenies, evaluating and recombining the best families. In order to increase the selection efficiency, it was proposed the use of the Genomic Recurrent Selection (SRG), which makes it possible to perform the evaluation and recombination phases simultaneously, besides providing a form of direct early selection. In the genomic selection dense panels of SNPs markers are used and several methods can be employed to obtain the selective prediction accuracy. The study was carried out with the UENF-14 population, obtained after the eighth cycle of Popcorn Recurrent Selection (SR). After recombination of the selected progenies, the seeds of the resulting population were sown, initiating the ninth SR cycle, which was used in this study. 200 individuals were sampled for genotyping, of which 98 individuals were self-pollinated for phenotypic evaluation of their S₁ progenies. Field trials were performed at two sites, under incomplete block design with three replicates. The parents of these progenies

were genotyped with a panel of 21,442 SNPs, and using the Plink software, three filters were performed, in the sequence: a) withdrawal of individuals with > 10% of missing data; b) discard of SNPs with > 5% of missing data; and c) SNPs discard with minor allele frequency -MAF <5%, resulting in 10,507 SNPs. The first chapter had as objective to obtain the selective accuracy and genetic gains for different selection strategies: FEN = estimates obtained only with the phenotypic information of 98 individuals; FEN + GEN = estimates obtained with the phenotypic and genotypic information of 98 individuals; and GEN = estimates obtained only with genotyping via SNPs markers. The 100 grains weight (P100), ear height (AE), grain yield (RG), popping expansion (CE), plant height (AP) and popcorn volume (VP) were evaluated. By the results obtained with the predictions using different selection intensities for the GEN strategy, it was possible to observe that the average annual genetic gain for the different traits was higher than the FEN and GEN + FEN strategies in 29.1% and 25.2% for 98 selection candidates; in 148.3% and 140.9% for 500; and 187.9% and 179.4% for 1,000 selection candidates, respectively. Therefore, the use of recurrent genomic selection may represent a high genetic gain, provided that: a) the phenotyping is performed accurately; b) the intensity of selection is explored with the genotyping of several individuals, increasing the number of candidates for selection, and c) using genomic selection for early selection in SR. In the second chapter the objective was to obtain the selective accuracy using the following methods: GBLUP, Bayes A, Bayes B, Bayes C π , Bayes Lasso and RKHS, define the best method and after this, define the best density of SNPs markers. Grain yield (RG), popping Volume (VP) and popping expansion (CE) were evaluated. The GBLUP method presented similar results to the others and was chosen for the next step, due to its execution simplicity, shorter time demanded in the analysis and greater numerical precision for the CE trait. From this point on, several scenarios with reduced panel SNPs were tested using the R software. In the different SNPs density scenarios, the results demonstrated that the use of ~ 4,800 SNPs presented selective accuracy similar to the use of 10,507 SNPs. It is concluded, then, that a considerable reduction of costs can be obtained by using a reduced panel of SNPs and that the GBLUP method should be used.

Keywords: genomic selection; accuracy; SNPs markers; popcorn.

1. INTRODUÇÃO

O milho-pipoca (*Zea mays* L.) é uma cultura de elevada rentabilidade entre os denominados “milhos especiais”, com influência positiva em setores da economia nacional e forte aceitação popular (Aguiar et al. 2008; Mendes de Paula et al., 2010; Rangel et al., 2011; Amaral Junior et al., 2013). Conforme o Anuário Brasileiro de Milho (2018), o Estado do Mato Grosso é considerado o principal produtor da cultura, com estimativa de 252 mil toneladas para o ano de 2018; porém, o país não é autossuficiente na produção da cultura para atender a demanda do mercado interno (Kist et al., 2018). Esta situação tem sido revertida por importantes avanços alcançados pelos programas de melhoramento (Arnhold et al., 2008; Scapim et al., 2010; Silva et al., 2010; Amaral Junior et al., 2013; Pena et al., 2016). O Brasil importou 12 mil toneladas de milho-pipoca em 2011, 13 mil toneladas em 2012, 6 mil toneladas em 2013 e 2 mil toneladas em 2014, revelando gradual decréscimo na importação (Senasa, 2018).

A Universidade Estadual do Norte Fluminense Darcy Ribeiro (UENF) atua no melhoramento da cultura desde 1998, com o intento de desenvolver novas cultivares de pipoca com características agronômicas favoráveis e alta capacidade de expansão (Pereira e Amaral Junior, 2001; Amaral Júnior et al., 2010; Freitas et al., 2014; Guimarães et al., 2016). Dentre as principais estratégias adotadas no melhoramento do milho-pipoca da UENF, destaca-se o melhoramento da população UNB via seleção recorrente, cujo procedimento permitiu a recomendação da própria

população melhorada como nova cultivar (Amaral Junior et al., 2013), além de possibilitar a contínua evolução da população para gerar linhagens elites.

O programa de melhoramento de milho-pipoca da UENF possui 14 cultivares registradas (MAPA, 2019); no entanto, a grande maioria dos registros no MAPA ainda pertence às empresas empacotadoras que estabelecem acesso restrito de uso com os produtores parceiros. Por conseguinte, o desenvolvimento de programas de melhoramento de milho-pipoca visando à obtenção de cultivares melhoradas são indispensáveis. Cabe ressaltar que a principal restrição ao uso dessas cultivares pelos pequenos produtores é a falta de sementes disponíveis para comercialização. Posto isso, foi criada em 2018 a 'Rio Norte Sementes', uma empresa de base tecnológica, iniciada em parceria com a Universidade Estadual do Norte Fluminense Darcy Ribeiro – UENF, com a qual será realizada a comercialização das sementes melhoradas.

A Seleção Recorrente (SR) é um dos métodos de melhoramento mais exequíveis para a obtenção de variedades de milho-pipoca, porém, é um procedimento trabalhoso e que demanda muito tempo, pois a cada ciclo de seleção são realizadas três etapas: obtenção das progênies, avaliação e recombinação das famílias superiores. O tempo necessário para completar um ciclo de SR pode requerer de 2,0 a 2,5 anos, o que reduz muito os ganhos genéticos anuais (Borém et al., 2017).

Com o intuito de reduzir o tempo necessário para a realização dos ciclos de seleção, a Seleção Genômica (GS), proposta por Meuwissen et al. (2001), é opção exequível para tanto. Embora a GS venha demonstrando ser uma importante ferramenta para aumentar os ganhos por seleção por unidade de tempo em dados simulados (Valente et al., 2016) e também para várias espécies de importância econômica, ainda não há trabalhos em populações reais de melhoramento com milho-pipoca visando a avaliar os ganhos genéticos para características de interesse agrônomo e a viabilidade econômica dessa técnica em comparação com as técnicas tradicionais de melhoramento genético.

Considerando-se que a UENF desenvolve um Programa de Melhoramento de Milho-Pipoca, utilizando a seleção recorrente, associar a este Programa a potencialidade da técnica da GS torna-se de relevância não apenas para o próprio Programa de Melhoramento da UENF, como também para a comunidade científica, que será beneficiada com as resultantes de pesquisa sobre a validade da técnica

SRG. Logo, a aplicação da técnica de SRG no milho-pipoca pode vir a contribuir pela rapidez na seleção de material genético mais adaptado às regiões Norte e Noroeste Fluminense.

2. OBJETIVOS

2.1. OBJETIVO GERAL

Utilizar a Seleção Recorrente Genômica na cultura milho-pipoca a fim de selecionar de forma otimizada as melhores progênies/famílias do Programa de Melhoramento da UENF.

2.2. OBJETIVOS ESPECÍFICOS

a) obter as acurácias seletivas para diferentes estratégias de seleção e os respectivos ganhos genéticos utilizando variadas intensidades de seleção, buscando, assim, averiguar a eficiência do uso da seleção recorrente genômica no melhoramento do milho-pipoca;

b) estimar as acurácias seletivas utilizando os seguintes métodos: GBLUP, Bayes A, Bayes B, Bayes Cp, Bayes Lasso e RKHS; e definir o método mais eficiente; e

c) testar diferentes densidades de marcadores de SNPs visando a verificar se com menor densidade de marcadores podem-se encontrar valores satisfatórios de acurácia, no intento de reduzir o custo do procedimento.

3. CAPÍTULOS

3.1. SELEÇÃO GENÔMICA PROPORCIONANDO MAIOR GANHO GENÉTICO ANUAL NO MELHORAMENTO POPULACIONAL DE MILHO-PIPOCA

3.1.1. INTRODUÇÃO

A Seleção Recorrente (SR) aumenta as frequências de alelos favoráveis para as características de interesse comercial (Hallauer et al., 2010, Borém et al., 2017) que no caso da cultura do milho-pipoca são a capacidade de expansão e o rendimento de grãos; além disso, conserva a variabilidade genética de forma a possibilitar novos ciclos de melhoramento. No entanto, é um procedimento trabalhoso e que demanda muito tempo. As etapas de avaliação e recombinação não são realizadas em nível de indivíduo, sendo, para tanto, utilizadas famílias oriundas das progênes consideradas superiores. Isto faz com que os ganhos com a seleção por ciclo sejam baixos (Borém et al., 2017).

Com a utilização da Seleção Recorrente Genômica (SRG) – procedimento que utiliza os princípios da Seleção Genômica (GS) na seleção recorrente – pode-se diminuir o tempo demandado para cada ciclo de seleção, pois a SRG possibilita realizar as fases de avaliação e de recombinação, simultaneamente. Esse método proporciona uma forma de seleção precoce direta, ou seja, atua prematuramente

sobre os genes expressos na idade adulta (Resende et al., 2010; Fritsche Neto et al., 2012). Por conseguinte, diminui-se o tempo necessário para obtenção de cada ciclo de seleção para apenas uma safra.

Buscando-se na literatura, ainda não há trabalhos com GS em populações reais de melhoramento com milho-pipoca visando a avaliar os ganhos genéticos para características de interesse agrônomo e a viabilidade dessa técnica em comparação com as técnicas tradicionais de melhoramento genético. No entanto, a GS tem demonstrado ser uma ferramenta eficiente para aumentar os ganhos de seleção por unidade de tempo, sendo utilizada na predição de efeitos de capacidade geral de combinação de 'testcrosses' e valores genotípicos de linhagens duplo-haplóides de milho (Albrecht et al., 2014); no melhoramento de alógamas (Yabe et al., 2013), na predição de valores genotípicos de híbridos simples de milho e arroz (Xu et al., 2014; Zhao et al., 2015).

Ademais, a GS, em alguns casos, tem sido mais eficiente por unidade de tempo do que a seleção fenotípica ou MAS (Heffner et al., 2010). Inobstante, outras pesquisas também têm comprovado que a GS é eficiente e pode ser utilizada no melhoramento vegetal (Daetwyler et al., 2013; De Los Campos et al., 2013; Valente et al., 2016).

Isto posto, o objetivo do presente estudo foi obter as acurácias seletivas para diferentes estratégias de seleção e os respectivos ganhos genéticos utilizando variadas intensidades de seleção, buscando, assim, averiguar a eficiência do uso da seleção recorrente genômica no melhoramento do milho-pipoca.

3.1.2. REVISÃO

3.1.2.1. MELHORAMENTO DO MILHO-PIPOCA NA UENF

No melhoramento de milho-pipoca, os principais objetivos são aumentar a qualidade da pipoca, tais como textura e maciez, definidos pela capacidade de expansão e, simultaneamente, o rendimento de grãos (Zinsly e Machado, 1987; Galvão et al., 2000; Scapim et al., 2002; Santos et al., 2007; Rangel et al., 2008). Segundo alguns melhoristas de milho-pipoca, o valor mínimo de CE para o

lançamento de uma cultivar é de 30 mL.g⁻¹ (Sawazaki, 2001; Matta e Viana, 2003; Scapim et al., 2010). O programa de melhoramento de milho-pipoca da Uenf possui a cultivar comercial “UENF HS02” no mercado com CE de 35 mL.g⁻¹, a qual é comercializada devido a parceria com a Rio Norte Sementes.

Em ordem hierárquica de importância, seguem as características relacionadas ao rendimento de grãos, como tamanho da espiga e do grão, além das características relacionadas à resistência a doenças e pragas (Santos et al., 2007). De forma geral, as plantas de milho-pipoca são mais suscetíveis ao acamamento, a pragas e doenças, ao quebramento dos colmos e à podridão de grãos, necessitando de um cuidado especial durante a colheita e a secagem dos grãos, para evitar danos no pericarpo e ao endosperma (Sawazaki, 2001). Assim, pode-se afirmar que há um ‘quantum’ elevado de características a serem melhoradas na cultura e que a definição da melhor estratégia e o uso de diferentes métodos de melhoramento são necessários.

A seleção recorrente, como método de melhoramento intrapopulacional associado às ferramentas biométricas, pode auxiliar na seleção de fenótipos superiores (Hallauer et al., 2010). Borém et al., (2017) ressaltam que os métodos intrapopulacionais, em geral, são os mais comumente utilizados do que os interpopulacionais, pois são de mais fácil execução e aplicáveis à maioria das características agronômicas.

O programa de melhoramento do milho-pipoca da UENF iniciou o processo de seleção recorrente com a população UNB-2U, proveniente de dois ciclos de seleção massal de UNB-1 que, por sua vez, derivou-se de um composto indígena e que foi cruzado com a Variedade “Americana” cujas progênies foram retrocruzadas com uma variedade “Amarela”, assim designada por conter grãos de mesma cor (Pereira e Amaral Junior, 2001). Ao longo dos ciclos de SR na população UNB-2U, foram obtidos acréscimos consideráveis nas principais características de importância econômica da cultura (Tabela 1); para capacidade de expansão (CE), houve evolução de 19,25 para 33,90 mL.g⁻¹ e, para rendimento de grãos (RG), de 1.699,44 para 3.368,42 kg.ha⁻¹ (Daros et al., 2002; Guimarães et al., 2018).

Tabela 1- Ciclos de seleção recorrente na população UNB-2U.

Ciclo	Citação	Estrutura de Família	Progresso no	Progresso
			RG (%)	na CE (%)
1°	Daros et al. (2002)	Irmãos completos	4,69	10,39
2°	Daros et al. (2004)	Auto fecundação S ₁	26,95	17,80
3°	Santos et al. (2007)	Meios-irmãos	10,00	7,16
4°	Freitas et al. (2009)	Irmãos completos	10,55	8,50
5°	Rangel et al. (2011)	Irmãos completos	8,53	6,01
6°	Ribeiro et al. (2012)	Irmãos completos	15,30	10,97
7°	Freitas et al. (2014)	Irmãos completos	7,78	5,11
8°	Guimarães et al. (2016)	Irmãos completos	4,6	3,61

Ciclo: ciclo de seleção; Citação: iniciada pelo primeiro autor da publicação do artigo referente ao respectivo ciclo; RG: rendimento de grãos; e CE: capacidade de expansão.

3.1.2.2. SELEÇÃO GENÔMICA

O grande atrativo da genética molecular em benefício do melhoramento genético aplicado é a utilização direta das informações de DNA na seleção, o que permite eficiência seletiva, grande rapidez na obtenção de ganhos genéticos com a seleção, em comparação com a tradicional seleção baseada em dados fenotípicos (Resende et al., 2008). Lande e Thompson (1990) propuseram o emprego da Seleção Assistida por Marcadores (SAM), cuja técnica utiliza simultaneamente dados fenotípicos e de marcadores moleculares em ligação gênica a locos controladores de caracteres quantitativos (QTL). Para as características quantitativas, o que tem limitado o seu uso são: demanda de estabelecimento de associações marcadores-QTLs para cada família em avaliação, necessidade de explicar grande parte da variação genética de uma característica quantitativa, proporcionar somente superioridade em relação à seleção baseada em dados fenotípicos quando o tamanho de população avaliado e genotipado é muito grande (da ordem de 500 ou mais) (Resende, 2008).

A Seleção Genômica (GS) foi proposta por Meuwissen et al. (2001), objetivando contornar as limitações da SAM e consiste na seleção com base no mérito genético predito, utilizando informações de um painel denso de marcas. Com essa abordagem, vislumbra-se tanto a seleção precoce, reduzindo

substancialmente o intervalo de geração do melhoramento (Goddard e Hayes 2009; Heffner et al., 2010; Resende Junior et al., 2012), quanto o aumento da acurácia, quando comparada com a predição fundamentada exclusivamente nas informações de genealogia (Almeida Filho et al., 2016). Bernardo e Yu (2007) foram alguns dos pioneiros a proporem a aplicabilidade da GS no melhoramento vegetal, onde obtiveram, via simulação de características quantitativas de milho, ganhos de seleção de 18 a 43 % maiores com o uso da GS em comparação ao uso da seleção recorrente assistida por marcadores moleculares.

Para aplicar a GS são empregados três tipos de população, sendo estas: EP, população de estimação ou treinamento; VP, população de validação; e SP, população de seleção. Dependendo do tipo, podem ser utilizadas diferentes estratégias e combinações destas populações para o emprego da GWS. Segundo Resende et al. (2010), tem-se utilizado comumente três populações diferentes ou uma só população para a estimação e validação.

Na prática da GWS as estimativas dos valores genéticos genômicos associadas a cada marcador ou alelo são utilizadas para o cálculo do valor genético genômico estimado (GEBV) de cada candidato à seleção. Assim, ao desenvolver a estimação dos modelos de GWS, a população deve ser fenotipada de modo preciso em condições de campo, e genotipada para estimação dos efeitos dos marcadores. A GS enfatiza a predição simultânea dos efeitos genéticos de milhares de marcadores dispersos no genoma, possibilitando capturar os efeitos dos locos tanto de pequenos quanto de grandes efeitos e explicar a quase totalidade da variação genética do caráter quantitativo. Em estudos de GS, é interessante definir se a predição será realizada utilizando as informações de '*pedigree*' e/ou as informações dos marcadores de DNA, pois desta forma pode-se optar pela substituição da matriz de parentesco baseada no '*pedigree*' pela matriz de parentesco estimada por marcadores. A vantagem do uso da matriz de parentesco estimada por marcadores é que computacionalmente é muito mais simples prever o GEBV de indivíduos diretamente, pois em geral o número de marcas é muito maior que o número de indivíduos (Crossa et al., 2010).

Em controvérsia com Jonas e Koning (2013), que consideram complexa a seleção genômica no melhoramento vegetal, em razão da diversidade de procedimentos seletivos, a técnica GS tem sido eficientemente utilizada por diversos autores (Daetwyler et al., 2013; De Los Campos et al., 2013; Zhao et al., 2013;

Albrecht et al., 2014; Xu et al., 2014; Cros et al., 2015; Zhao et al., 2015; Mafra et al., 2019).

3.1.3. MATERIAL E MÉTODOS

3.1.3.1. POPULAÇÃO DE ESTUDO, GENOTIPAGEM E FENOTIPAGEM

O estudo foi realizado com a população UENF-14, obtida após o oitavo ciclo de Seleção Recorrente (SR) de Milho-Pipoca. Após a recombinação das progênes selecionadas, foram semeadas as sementes da população resultante, iniciando o nono ciclo de SR, o qual foi utilizado neste trabalho. Foram amostrados 200 indivíduos para a genotipagem sendo que desses, 98 indivíduos foram autofecundados, para avaliação fenotípica de suas progênes S₁.

A partir das 200 amostras de DNA de plântulas coletadas, foi realizada a caracterização do polimorfismo no genoma entre indivíduos dessa população, utilizando a técnica 'Capture Seq', considerando 5.000 sondas bem distribuídas no genoma de referência do milho. Essa abordagem genômica, foi realizada em colaboração com a empresa 'Rapid Genomics' LLC, resultando em 21.442 SNPs. De posse das informações genotípicas, com o software Plink (Purcell e Chang, 2015), foram realizados três filtros, na sequência: a) retirada de indivíduos com > 10 % de dados perdidos; b) descarte de SNPs com > 5 % de dados perdidos; e c) descarte de SNPs com MAF (frequência do alelo menos comum) < 5 %, resultando em 196 indivíduos com 10.507 SNPs. Desses indivíduos filtrados, 98 foram utilizados no ensaio de progênes S₁.

Com os 196 indivíduos e 10.507 SNPs, foram realizadas caracterizações genômicas da população, estimando o desequilíbrio de ligação (LD), utilizando a estatística r^2 por meio do 'software' Plink (Purcell e Chang, 2015). Com os resultados de LD, foi ajustada uma curva descritiva do decaimento, conforme Hill e Weir (1988), usando a função **nlm** da linguagem R. Além disso, foi calculado o coeficiente de parentesco dos indivíduos, obtendo a Matriz de parentesco genético, usando o pacote rrBLUP (Endelman, 2011), contudo, somando um pequeno valor de 10^{-2} na

diagonal, para melhorar a estabilidade numérica (Meuwissen et al., 2011; Ertl et al., 2014).

Com as famílias S_1 obtidas, foram delineados dois ensaios, sendo um no Colégio Estadual Agrícola Antônio Sarlo, em Campos dos Goytacazes, região Norte do Estado do Rio de Janeiro, situado a $21^\circ 45'$ de latitude sul e $41^\circ 20'$ W de longitude e a 11 m de altitude, classificado como tropical chuvoso, clima de bosque com uma precipitação média anual de 1.023 mm, evapotranspiração potencial de 1.601 mm anuais e temperatura média anual de 23° C, e o outro experimento na Estação Experimental da PESAGRO-RIO de Itaocara, localizada na Região Noroeste Fluminense, situada a $21^\circ 39' 12''$ de latitude sul e $42^\circ 04' 36''$ W de longitude e a 60 m de altitude, com temperatura média anual de $22,5^\circ$ C e precipitação média anual de 1.041 mm, ambos implantados em agosto de 2016.

O delineamento experimental utilizado foi blocos incompletos, com três repetições. Para tanto, cada família foi semeada em linha de 5,00 m de comprimento, com espaçamento entre linhas de 0,90 m e entre plantas de 0,20 m, com três sementes por cova, à profundidade de 0,05 m. Aos 21 dias após a emergência, foi realizado o desbaste, deixando-se uma planta por cova, totalizando uma população de 60.000 plantas por hectare. A adubação no plantio foi realizada de acordo com análise de solo. A adubação de cobertura foi realizada cerca de 30 dias após o plantio. Os tratos culturais foram realizados conforme as necessidades da cultura.

As características fenotipadas foram: a) altura média da planta, que é a mensuração, em cm, a distância do nível do solo à inserção da folha bandeira, logo após o pendoamento, de seis plantas competitivas; b) altura média da espiga, que se refere à quantificação da distância, em cm, do nível do solo à base de inserção da primeira espiga em seis plantas por parcela; c) peso de 100 grãos em gramas; d) rendimento de grãos, expresso em $\text{kg}\cdot\text{ha}^{-1}$; e) capacidade de expansão, determinada em vasilha plástica sem óleo, com três repetições por parcela com peso de 30 g de grãos cada. O pipocamento foi realizado no microondas durante 2 minutos, após o qual, o volume da pipoca expandida foi mensurado em proveta de 2.000 mL e o resultado dividido pela massa inicial de 30 g de grãos e expresso pela relação mL g^{-1} ; e f) volume de pipoca expandido por hectare (VP), obtido pela multiplicação entre a produtividade média da parcela e a capacidade de expansão,

gerando o volume médio de pipoca expandido por hectare de plantio, expresso em $m^3 \cdot ha^{-1}$ ($VP = RG \times CE / 1000$).

3.1.3.2. MÉTODO ESTATÍSTICO APLICADO NA SELEÇÃO GENÔMICA

Neste trabalho foi utilizado um modelo de indivíduos, sendo realizada a substituição da matriz de parentesco baseada no ‘pedigree’ pela matriz de parentesco estimada por marcadores, cujo modelo é designado de ‘Genomic – Best Linear Unbiased Predictor’ (GBLUP). A vantagem do GBLUP consiste no fato de que computacionalmente é muito mais simples prever o ‘Genomic Estimated Breeding Value’ (GEBV) de indivíduos diretamente, pois em geral o número de marcas é muito maior que o número de indivíduos (Misztal e Legarra, 2016).

Os modelos GBLUP são baseados na matriz GRM, obtida com as informações do painel de marcadores (Misztal, 2016), sendo que, para obtenção dessa matriz é assumido que todo marcador com mesma MAF contribui igualmente para a variância genética. Assim, pode-se dizer que esse modelo é condizente com o modelo de herança infinitesimal, tradicionalmente assumido na genética quantitativa clássica, que pressupõe que as características são poligênicas com número de genes que tende ao infinito e todos contribuem igualmente com a variação fenotípica. Considerando as condições experimentais do presente estudo, o modelo GBLUP ajustado foi:

$$y = X\beta + Z_1b + Z_2u + Z_3(u \times l) + \varepsilon ,$$

em que: y é o vetor de informações fenotípicas de uma determinada característica; β é o vetor de efeitos fixos que incluem o intercepto do modelo, efeito de repetição dentro de local, efeito de local e co-variáveis como número de plantas na parcela, contado logo após o desbaste e umidade de grãos, apenas para as características P100, RG, CE e VP; b é o efeito aleatório de bloco incompleto dentro de repetição e local; u é o efeito aleatório do valor genético aditivo genômico (GEBV); $u \times l$ é o efeito aleatório da interação entre GEBV e local; e ε é o efeito do erro aleatório.

Para este modelo foi assumido que:

$$b \sim N(0, I_{(b)}\sigma_b^2); u \sim N(0, G\sigma_A^2); (u \times l) \sim N(0, G \otimes I_{(l)}\sigma_{AL}^2); \varepsilon \sim N(0, I_{(\varepsilon)}\sigma_\varepsilon^2) ,$$

em que: $I_{(b)}$, $I_{(l)}$ e $I_{(\varepsilon)}$ são matrizes identidade com ordem igual ao comprimento do vetor b , número de ambientes e número de observações, respectivamente; G é a

matriz GRM descrita anteriormente; \otimes é o produto de Kronecker; σ_b^2 , σ_A^2 e σ_{AL}^2 são os componentes de variância referentes aos blocos, variância aditiva e componente de variância da interação genética aditiva com o local, respectivamente. Os componentes de variância foram estimados via REML com o algoritmo AI utilizando pacote **asreml** (Butler, 2009) do ‘software’ R (R Core Team, 2013).

Também foram estimados componentes de correlação genética e ambiental, via modelos ‘*bi-trait*’ e ‘*bi-environment*’,

$$[y_i \ y_j] = [X_i \ 0 \ 0 \ X_j][\beta_i \ \beta_j] + [Z_{1i} \ 0 \ 0 \ Z_{1j}][b_i \ b_j] + [Z_{2i} \ 0 \ 0 \ Z_{2j}][u_i \ u_j] + [\varepsilon_i \ \varepsilon_j],$$

em que os subscritos **i** e **j** representam dois traços em mesmo ambiente, ou o mesmo traço nos dois ambientes, considerando-se neste modelo que:

$$[b_i \ b_j]' \sim N(0, \text{diag}\{\sigma_{b_i}^2, \sigma_{b_j}^2\} \otimes I_{(b)}); [u_i \ u_j]' \sim N(0, G_0 \otimes G); [\varepsilon_i \ \varepsilon_j]' \sim N(0, R_0 \otimes I_{(\varepsilon)}),$$

em que $G_0 = [\sigma_{a_i}^2 \ r_a \sigma_{a_i} \sigma_{a_j} \ \text{sim.} \ \sigma_{a_j}^2]$ e $R_0 = [\sigma_{\varepsilon_i}^2 \ r_\varepsilon \sigma_{\varepsilon_i} \sigma_{\varepsilon_j} \ \text{sim.} \ \sigma_{\varepsilon_j}^2]$. Os outros termos do modelo são análogos ao modelo ‘*uni-trait*’.

3.1.3.3. ESTIMATIVAS DA ACURÁCIA SELETIVA E GANHO DE SELEÇÃO

A seleção com uso da genotipagem via SNPs será eficiente somente com um aumento considerável na resposta à seleção por unidade de tempo quando comparado com a seleção baseada na média de famílias, posto que a comparação dessas médias é a base da seleção das melhores progênes ao longo de vários ciclos do programa de seleção recorrente da UENF-14.

Desta forma, para obter as estimativas da acurácia seletiva e do ganho de seleção foram utilizadas três diferentes estratégias, a saber: a) FEN= estimativas obtidas somente com as informações fenotípicas das 98 progênes; b) FEN + GEN= estimativas obtidas com as informações fenotípicas e genotípicas dos 98 progênes; e c) GEN= estimativas obtidas somente com a genotipagem via os marcadores SNPs; sendo o modelo GBLUP ajustado com 90 % dos indivíduos da população, enquanto a parte remanescente (10 %) tem o mérito genético predito considerando apenas as informações de marcadores. Esse processo de ajuste de modelo foi repetido dez vezes, sendo que em cada ciclo, um grupo diferente de indivíduos teve o mérito genético predito com o fenótipo ignorado. Esse processo é conhecido como validação cruzada ‘ten-fold’ (Resende Junior et al., 2012; Almeida Filho et al., 2016),

e nesse estudo foi utilizada para verificar a eficácia da seleção precoce, considerando um modelo previamente ajustado.

Com a estratégia GEN foram realizadas predições, empregando diferentes intensidades de seleção, utilizando de 98 até 1.000 candidatos à seleção e fixando o valor absoluto de 40 indivíduos selecionados. Após isso, foi realizada uma avaliação comparativa do ganho genético anual obtido via as estratégias GEN, FEN e GEN + FEN. A acurácia foi estimada conforme Gilmour et al., (2015), sendo considerada neste trabalho apenas as estimativas dos 98 genitores que passaram no filtro para taxa de dados de SNPs perdidos.

Para o ganho genético foi utilizada a seguinte fórmula:

$$\Delta G = k r \sigma_A / T$$

Em que: **k** = diferencial de seleção padronizado, tendo sido fixado o número de 40 indivíduos selecionados; **r** = acurácia seletiva, determinada pela acurácia preditiva estimada por cada método; **σ_A** = raiz quadrada da variância aditiva, estimada considerando as informações genóticas e fenóticas; e **T** = tempo para completar um ciclo; sendo que para as estratégias FEN e GEN + FEN é necessário fenotipar as progênes e a Seleção Recorrente tem ciclo com tamanho normal de três safras (T = 1,5 ano em condições tropicais), enquanto no caso da estratégia GEN, em que prescinde a etapa de fenotipagem, a seleção é feita em nível de planta e não de progênes, e como a seleção é realizada antes da floração, os indivíduos selecionados são diretamente recombinados e, assim, um ciclo de SR com modelo previamente ajustado demanda uma safra (T = 0,5 ano em condições tropicais). Foram utilizados os tempos de 0,5 ano para a estratégia GEN e 1,5 anos para a estratégia GEN + FEN, considerando-se a possibilidade de realizar safras consecutivas de seleção recorrente. Em muitos cenários o ambiente exige um tempo de 2 anos para realizar o ciclo completo, devido à impossibilidade de realizar uma safra na estação fria.

3.1.4. RESULTADOS E DISCUSSÃO

3.1.4.1 POPULAÇÃO DE ESTUDO E PARÂMETROS GENÉTICOS

Com intuito de conhecer melhor algumas características genéticas da população de estudo, foram estimados alguns parâmetros com o painel de SNPs disponível. Após a realização dos filtros para dados perdidos, pouco mais de 10 % dos SNPs apresentaram alelos fixados na população, enquanto ~25 % se revelaram polimórficos; contudo com baixa MAF ($<0,05$). Portanto, uma quantidade satisfatória de ~65 % das marcas bem distribuídas no genoma possuem seus alelos segregando com elevada frequência na população de estudo (Figura 1a). O r^2 entre SNPs adjacentes foi em média ~0,36 (SD=0,36), sendo a metade do decaimento estimada em ~107.54 Kb (Figura 1b). O coeficiente de parentesco obtido com os marcadores está distribuído em torno de zero (Figura 1c), condizente com uma população oriunda de recombinação aleatória de plantas; porém, o coeficiente de endogamia de ~89 % desses indivíduos, resultantes de oito ciclos de SR prévios, foi estimado entre 0,1 e 0,3 (Figura 1d).

O componente de variância genética foi significativo em todos os caracteres avaliados para $p < 0,01$ (Tabela 2). Estas diferenças genéticas significativas observadas indicam existência de variabilidade genética, o que possibilita a prática da seleção com consequentes ganhos genéticos dentre os genótipos em estudo. O ganho genético no longo prazo depende, fundamentalmente, da variabilidade genética potencial, que é mantida por meio dos ciclos seletivos e é liberada através da recombinação, ao final de cada ciclo. Assim, o estabelecimento das populações de melhoramento deve ser analisado com base na variabilidade genética (Robertson, 1960; Terra et al., 2011).

A variância da interação genótipo x local foi significativa somente para os caracteres RG ($p < 0,01$) e VP ($p < 0,01$). Isto ocorre quando há respostas diferenciadas dos genótipos testados em diferentes ambientes. Porém, inferir se a interação foi significativa, é apenas uma abordagem qualitativa, sendo que apenas com a correlação genética entre os ambientes, é possível investigar a resposta correlacionada (Kashiani e Saleh, 2010). Para a característica RG, de correlação genética elevada, a seleção em um ambiente acarretará em resposta positiva no outro. Como a correlação genética entre os ambientes foi positiva, a seleção com base no valor aditivo geral, isto é, considerando os dois ambientes, vai proporcionar resposta seletiva positiva para ambos os ambientes.

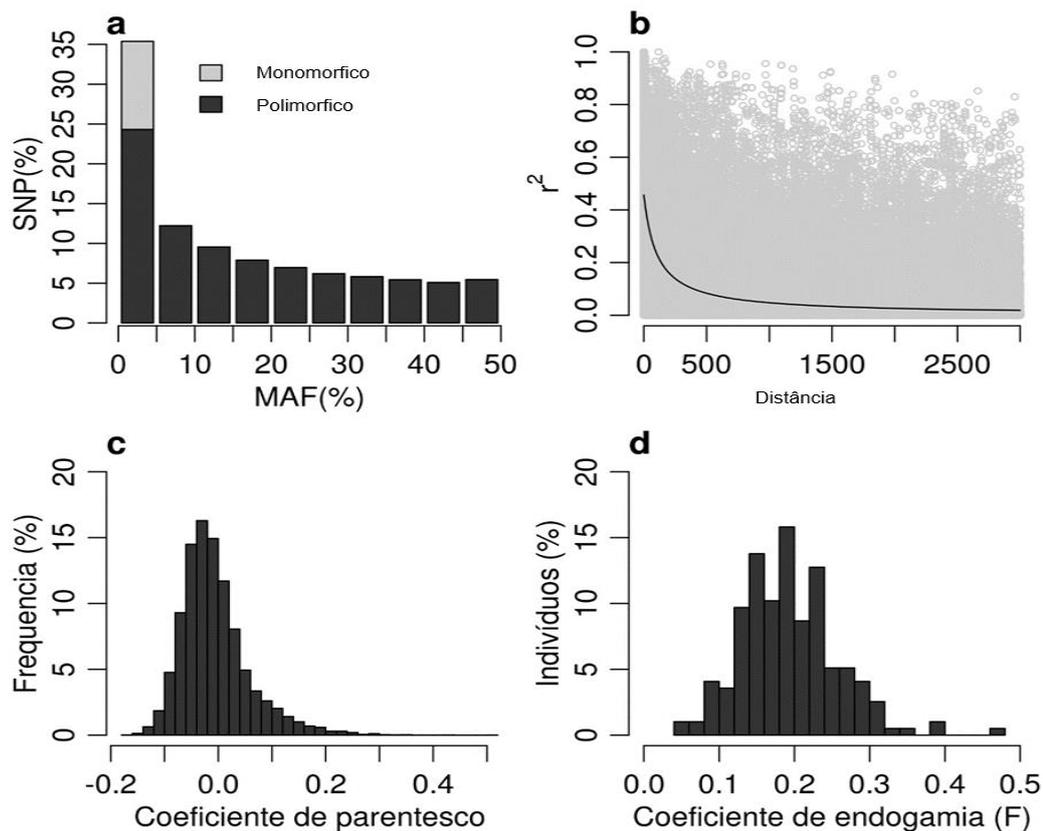


Figura 1 - Aspectos genéticos da população de estudo. a) Distribuição de frequências alélicas; b) Desequilíbrio de ligação; c) Distribuição do coeficiente de parentesco entre os indivíduos; d) Distribuição do coeficiente de endogamia dos indivíduos. A figura **a** foi gerada com um painel de 16.258 SNPs com menos de 5 % de dados perdidos e as figuras **b**, **c** e **d** foram criadas com 10.507 SNPs da figura **a** com MAF > 5 %. Em todas as figuras foram considerados 196 indivíduos com taxa de dados perdidos menor que 10 %.

Os coeficientes de variação experimental apresentaram valores aceitáveis (Tabela 2) para as características em estudo (Fritsche-Neto et al., 2012). No presente estudo, as herdabilidades com base na média, apresentaram estimativas de média à alta magnitude para a maioria das características, sendo inferior a 0,5 apenas para as características VP e superiores a 0,8 para P100, AE e AP (Tabela 2). A herdabilidade se trata de um dos parâmetros genéticos de maior importância dentro do melhoramento genético, por ter atribuição preditiva, expressando a confiança do valor fenotípico como um guia para o valor genético (Borém et al., 2017). Portanto, é um parâmetro-chave na escolha de estratégias de melhoramento, uma vez que influencia diretamente na tomada de decisão entre quais técnicas de seleção serão utilizadas, tornando mais eficiente o processo de obtenção de novas cultivares (Hallauer et al., 2010).

Para os caracteres de maior importância agrônômica – CE e RG –, obtiveram-se herbabilidades de 0,62 e 0,51, respectivamente (Tabela 2). Amaral Junior et al. (2016) também observaram menor influência ambiental na manifestação de CE, quando comparada com RG. A maior herdabilidade de CE pode ser explicada ainda pelo fato do caráter ser governado predominantemente por genes aditivos (Larish e Brewbaker, 1999; Pereira e Amaral Junior, 2001; Santos et al., 2007; Rangel et al., 2008; Schwantes et al., 2017). A baixa herdabilidade (0,43) obtida com o caráter VP, pode ser explicada pelo fato desta ser uma combinação de duas características RG e CE, agregando também a influência ambiental na manifestação do caráter de duas características poligênicas. Segundo Heslot et al. (2015) a seleção genômica proporcionará ganhos genéticos relevantes ao melhoramento vegetal, mesmo em caracteres de baixa herdabilidade, uma vez que ela diminui a duração do ciclo de seleção. As elevadas herdabilidades (>0,80) obtidas para os caracteres P100, AE e AP, indicam a confiança da transmissão das características a futuras gerações, implicando diretamente nos ganhos por seleção (Hallauer et al., 2010).

Pode-se notar que a herdabilidade apresentou alta correlação com a acurácia para as diferentes estratégias ($GEN = 0,99$, $FEN + GEN = 0,93$ e $FEN = 0,94$), sendo que maior herdabilidade representou maiores valores de acurácia. A alta correlação da herdabilidade com a acurácia encontrada concorda com resultados obtidos por Combs e Bernardo (2013), Lian et al. (2014) e Zhang et al. (2017). Segundo Zhang et al. (2017), quando os melhoristas projetam os experimentos de fenotipagem da GS, estes devem considerar que a herdabilidade das características de interesse na população de treinamento deve ser alta para alcançar bons valores de acurácia preditiva e, para tanto, devem aumentar o número de locais e repetições.

Tabela 2 - Estimativas da acurácia seletiva considerando apenas o fenótipo (FEN), considerando fenótipo e marcas (FEN + GEN) simultaneamente e considerando apenas marcas (GEN); componentes de variância aditiva (Va), da interação dos efeitos aditivos x locais (Val), e herdabilidade média para as características avaliadas.

Característica	Média	CV (%)	Acurácia Média			h ²	Componentes de Variância	
			FEN	FEN + GEN	GEN		Va	Val
P100 (g)	13,26	6,35	0,87	0,89	0,38	0,84	0,70(<0,01)	0,03 (0,42)
AE (cm)	99,04	8,27	0,89	0,90	0,39	0,87	89,74(<0,01)	5,07 (0,18)
RG (Kg/ha)	2546,08	24,11	0,74	0,76	0,31	0,51	88134,78(<0,01)	46082,10 (0,06)
CE (mL/g)	28,23	12,39	0,74	0,75	0,34	0,62	3,68 (<0,01)	0,42 (0,45)
AP (cm)	185,78	5,68	0,9	0,91	0,39	0,89	150,26 (<0,01)	0,23 (0,97)
VP (m ³ /ha)	72,3	27,61	0,74	0,76	0,3	0,46	87,58 (<0,01)	78,23 (0,02)

Peso de 100 grãos (P100), altura de espiga (AE), rendimento de grãos (RG), capacidade de expansão (CE), altura de planta (AP) e volume de pipoca (VP). # Os valores entre parêntesis são os p-valores, obtidos pelo teste LRT.

3.1.4.2. ACURÁCIA NA SELEÇÃO GENÔMICA

A acurácia na seleção genômica é determinada por meio da correlação entre as estimativas do valor genético genômico estimado e os valores genéticos obtidos via fenotipagem (Meuwissen et al., 2001), sendo que a abordagem da validação cruzada é geralmente empregada para avaliar a acurácia da seleção genômica (Erbe et al., 2010). Estas acurácias obtidas na validação são utilizadas para predição dos futuros fenótipos candidatos à seleção (Bernardo e Yu, 2007; Resende, 2008; Crossa et al., 2017).

Na seleção recorrente convencional, a seleção é realizada com base apenas nas informações fenotípicas; neste estudo, ao incluir a informação do conjunto de 10.507 marcadores SNPs, ocorreu um pequeno aumento na acurácia para maioria dos caracteres, em que em média o modelo GEN + FEN foi ~2,5 % mais acurado que o modelo tradicional, que desconsidera os marcadores (FEN). Esta superioridade da acurácia para a maioria das características avaliadas em relação à acurácia da estratégia FEN, deveu-se ao incremento ocasionado pelo uso da matriz de parentesco realizada; ou seja, baseada nas marcas que os indivíduos apresentam e não como a tradicional, com base em esperança matemática (Liu et al., 2015).

As acurácias obtidas apenas com a estratégia GEN variaram de 0,25 a 0,39, sendo estes valores bem inferiores aos obtidos com as estratégias FEN e GEN + FEN (Tabela 2). Da mesma forma, outros trabalhos retratam que a predição do mérito genético, considerando apenas as informações de parentesco – baseado em genealogia e/ou marcadores –, mostraram que a acurácia é menor, na ausência das informações fenotípicas (Muranty et al., 2015; Lenz et al., 2017; Norman et al., 2017). Dessa forma, naturalmente, para se aplicar a GS no melhoramento, devem-se explorar outros componentes do ganho genético, como intensidade de seleção e a redução do intervalo entre gerações, pois a fenotipagem realizada de forma precisa, por si só, proporciona razoáveis acurácias seletivas.

Os caracteres relacionados ao desenvolvimento da cultura, como AP e AE apresentaram os maiores valores de acurácia para as diferentes estratégias. O fato da característica relacionada ao desenvolvimento da cultura, AP, apresentar valores de acurácia superiores ao importante caráter agrônômico RG para as diferentes estratégias, concorda com resultados encontrados por Zhang et al. (2017), em milho comum. Para as características de maior interesse agrônômico na cultura do milho-pipoca (RG, CE e VP), obtiveram-se valores de acurácia superiores a 0,74 com as estratégias GEN + FEN e FEN. Além disso, a característica agrônômica secundária P100 também apresentou valores elevados de acurácia e herdabilidade. De acordo com Resende et al. (2010), valores de acurácia iguais ou superiores a 0,70 são desejáveis; assim, os valores superiores a 0,74 encontrados para as características em estudo demonstram que podem ser obtidos ganhos genéticos satisfatórios utilizando a seleção genômica.

Valente et al. (2016), em estudo simulado com milho-pipoca identificaram acurácia de 0,55 para herdabilidade de 0,30; e acurácia de 0,84 para herdabilidade de 0,70, ambas para a característica CE. No presente estudo, com dados reais de CE foram obtidos os valores de 0,75 de acurácia para herdabilidade de 0,62, considerando-se a estratégia GEN + FEN. Pode-se inferir, desta forma, que os resultados são consonantes, dada a similaridade nas proporções encontradas.

3.1.4.3. CORRELAÇÕES GENÉTICA E AMBIENTAL E GANHO GENÉTICO

Hallauer et al. (2010) relataram a importância dos coeficientes de correlação simples, uma vez que, com esses, pode-se quantificar o grau de associação

genético e não-genético entre dois ou mais caracteres. A correlação genética pode ser devida à pleiotropia ou à ligação gênica (em uma população em desequilíbrio de ligação) entre genes que são responsáveis por duas características. Alguns genes podem aumentar o valor fenotípico de duas características, causando uma correlação positiva, enquanto outros genes podem aumentar uma e reduzir a outra, causando correlação negativa (Falconer, 1981).

No presente estudo, as correlações genéticas entre os ambientes foram elevadas para a maioria das características, apresentando valores próximos a 1, sendo exceção a característica VP que teve correlação de aproximadamente 0,5 para com as demais. No presente estudo, as correlações entre os ambientes foram elevadas para a maioria das características, destacando-se RG e VP, e AE e AP. Segundo Falconer (1981), o ambiente torna-se causa de correlações quando duas características são influenciadas pelas mesmas variações de condições ambientais, em que valores negativos dessa correlação evidenciam que o ambiente favoreceu uma característica em detrimento da outra, e quando ocorrem valores positivos, indica que ambas foram beneficiadas ou prejudicadas pelas mesmas causas de variações ambientais. Com relação à correlação residual, esta se mostrou praticamente nula para com todas as características ao se considerar ambos os ambientes (Figura 2).

Dentre as correlações genéticas observadas destacam-se os pares de caracteres AP e AE; RG e VP; os quais se mostraram correlacionados de forma elevada ($>0,95$) e positiva em ambos os ambientes (Figura 2). Naturalmente as plantas com maiores alturas de espigas tendem ser as plantas maiores (Cabral et al., 2016). Segundo Cruz et al. (2012) essas associações podem possibilitar ganhos indiretos por seleção em caracteres correlacionados.

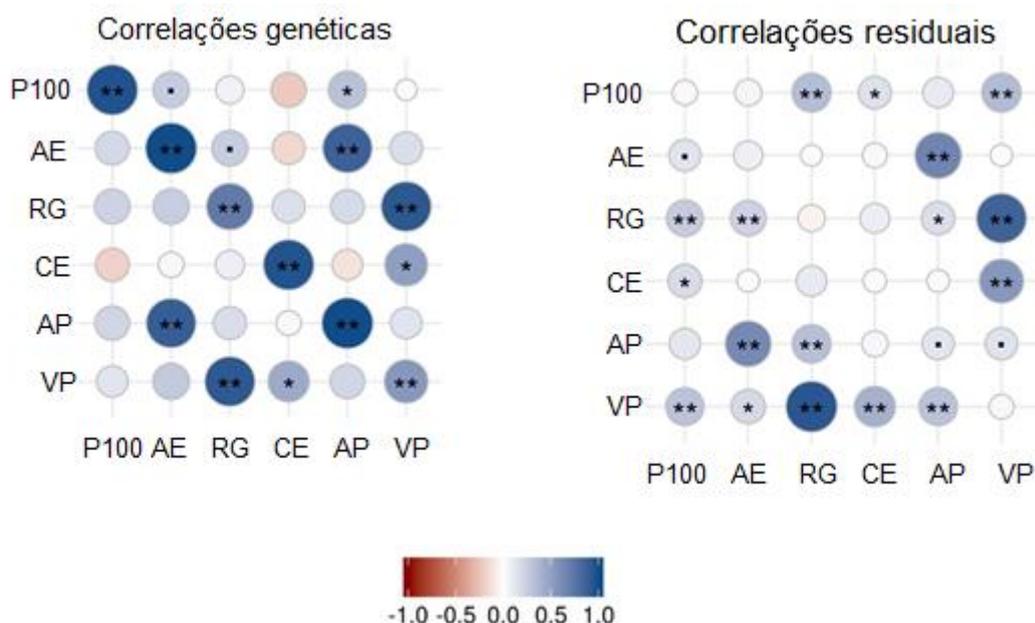


Figura 2 - Gráficos das correlações genéticas e residuais obtidas para as características peso de 100 grãos (P100), altura de espiga (AE), rendimento de grãos (RG), capacidade de expansão (CE), altura de planta (AP), e volume de pipoca (VP). Correlações entre as características, estimadas em Campos dos Goytacazes e Itaocara estão acima e abaixo da diagonal, respectivamente; na diagonal estão as correlações de cada característica entre os locais. **, * e dot ".": significam correlações estatisticamente diferente de zero a 1, 5 e 10 % de significância.

Ganho genético é o montante de acréscimo em performance que é alcançado via os programas de melhoramento por meio de consecutivos ciclos de seleção (Crossa et al., 2017). O ganho de seleção depende do diferencial de seleção, que por sua vez é a diferença entre a média do grupo selecionado e a média da população original. Portanto, em processo de seleção, quanto maior for a pressão de seleção, maior será esse diferencial e, conseqüentemente, o progresso genético (Borém et al., 2017).

A partir dos resultados obtidos com as predições utilizando diferentes intensidades de seleção para a estratégia GEN foi possível observar que o ganho genético anual médio para as diferentes características foi superior às estratégias FEN e GEN + FEN em 29,1 % e 25,2 % para 98 candidatos à seleção; em 148,3 % e 140,9 % para 500; e atingiu 187,9 % e 179,4 % para 1.000 candidatos à seleção, respectivamente. A superioridade da estratégia GEN em relação à FEN e GEN + FEN, concorda com resultados obtidos por Wong e Bernardo (2008), Jannink et al. (2010) e Yabe et al. (2013), os quais indicam uma potencial superioridade da GS

sobre seleção fenotípica. No entanto, isso depende de um modelo de predição ajustado, para ser usado em um ou mais ciclos de SR sem ser recalibrado. De acordo com Muller et al. (2017), o sucesso desta estratégia depende em grande parte da persistência de valores significativos de acurácia, quando os candidatos a seleção são separados por um ou mais ciclos da geração de treinamento do modelo.

De acordo com Jannink et al. (2010), não é possível afirmar por quantos ciclos o modelo será efetivo, pois mudanças no padrão LD, frequências alélicas e a perda de polimorfismos são imprevisíveis. De acordo com Fritsche-Neto et al. (2012), é especulado que pode-se usar a GS de um a três ciclos de SR sem a avaliação fenotípica; no entanto, essa afirmativa ainda carece de ser melhor avaliada. Já Combs e Bernardo (2013) realizaram cinco ciclos de GS e observaram que os ganhos genéticos realizados no rendimento de grãos de milho geralmente concordavam com o nível previsto, embora os ganhos após o primeiro ciclo fossem instáveis.

Observando-se os resultados, verifica-se que a estratégia FEN + GEN apresentou 3 % de ganho médio anual em relação à estratégia FEN (Figura 3). Essa superioridade deve-se ao fato de a estratégia FEN ignorar as informações de marcadores para obtenção da matriz de parentesco. Na população UENF-14 a matriz de parentesco é a própria identidade ($G=I$), pois é uma população não estruturada, devido à recombinação realizada, com as progênies selecionadas do ciclo de SR, imediatamente anterior a esse trabalho. De acordo com Liu et al. (2017), quanto mais próxima a relação genética real entre a população treinamento e a de validação, menos marcadores serão necessários para atingir uma boa precisão de previsão.

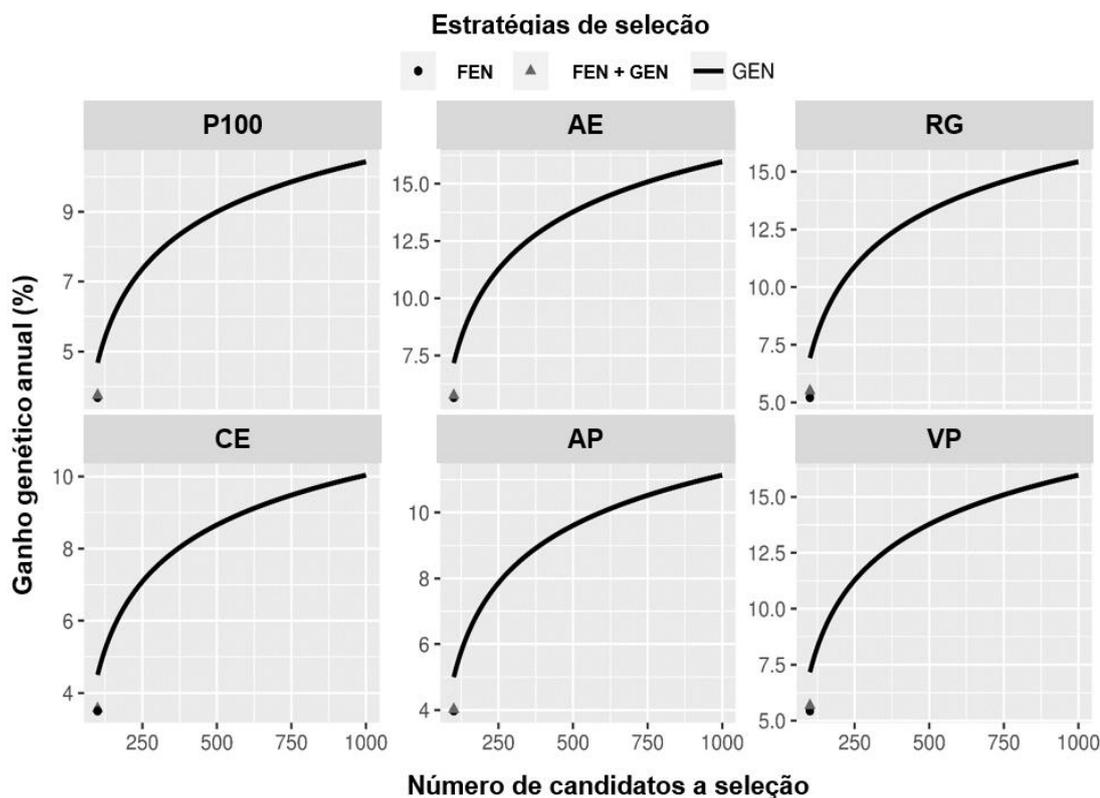


Figura 3 - Ganho genético anual para as diferentes estratégias utilizadas. Ganho genético anual para as diferentes estratégias utilizadas nas características peso de 100 grãos (P100), altura de espiga (AE), rendimento de grãos (RG), capacidade de expansão (CE), altura de planta (AP), e volume de pipoca (VP). As estratégias de seleção consideradas foram baseadas exclusivamente nos fenótipos (FEN), considerando tanto os fenótipos, quanto os marcadores (FEN + GEN) e finalmente, considerando apenas as informações dos SNPs.

As características de maior importância agrônoma – RG e CE – expressaram ganhos anuais significativos para a estratégia GEN, chegando a ser 32,8 % e 28,2 % superiores à estratégia FEN com 98 candidatos a seleção, e a 155,4 % e 147,1 % superiores com 500 candidatos a seleção, respectivamente, para a mesma estratégia FEN com 100 candidatos à seleção. Pode-se observar na estratégia GEN que o ganho médio anual apresentou um aumento de 92,4 % com até 500 candidatos à seleção, e que após isto seguiu ocorrendo um acréscimo no ganho; no entanto, em menor magnitude, pois de 500 para 1.000 candidatos a seleção, o ganho foi de 15,9 % (Figura 3). Isto posto, os ganhos anuais preditos nas características RG e CE para a estratégia GEN foram muito superiores aos obtidos em trabalhos em ciclos anteriores de SR nesta população, em que a seleção foi realizada exclusivamente com avaliação de médias fenotípicas de famílias (Daros

et al., 2002; Daros et al., 2004; Santos et al., 2007; Freitas Júnior et al., 2009; Rangel et al., 2011; Ribeiro et al., 2012; Freitas et al., 2014; Guimarães et al., 2016).

3.1.5. CONCLUSÕES

O uso da seleção recorrente genômica proporciona elevado ganho genético, desde que: a) a fenotipagem seja realizada de modo preciso, b) seja explorada a intensidade de seleção com a genotipagem de vários indivíduos, aumentando o número de candidatos à seleção; e c) utilize a GS para seleção precoce na SR, reduzindo assim, o intervalo entre os ciclos, por prescindir do teste de progênies. Contudo, mais estudos são necessários para avaliar até quantos ciclos de SR subsequentes um modelo de predição genômica pode ser aplicado para se obter GEBV com acurácias razoáveis.

3.2. ANÁLISE COMPARATIVA DE ALGORITMOS BAYESIANOS E FREQUENTISTAS NA EFICIÊNCIA DA SELEÇÃO RECORRENTE GENÔMICA EM MILHO-PIPOCA E OTIMIZAÇÃO DE DENSIDADES DE MARCADORES SNPs

3.2.1. INTRODUÇÃO

A Seleção Recorrente Genômica (SRG) – procedimento que utiliza os princípios da Seleção Genômica (GS) na seleção recorrente – diminui o tempo demandado para cada ciclo de seleção recorrente, pois possibilita realizar as fases de avaliação e de recombinação, simultaneamente. Esse método proporciona uma forma de seleção precoce direta, ou seja, atua prematuramente sobre os genes expressos na idade adulta (Resende et al., 2010; Fritsche Neto et al., 2012). Por conseguinte, diminui-se o tempo necessário para obtenção de cada ciclo de seleção para apenas uma safra.

A seleção genômica (GS), proposta por Meuwissen et al. (2001), é uma abordagem que consiste na predição do potencial genético de indivíduos, usando informações de grande número de marcadores amplamente distribuídos ao longo do genoma. Essa predição é realizada por meio de modelos, onde a caracterização genômica resultante da genotipagem é utilizada como fonte de variação explicativa e o fenótipo de uma dada característica de interesse é utilizado como variável resposta. Esse modelo de predição genômica é ajustado na população de treinamento (PT) – onde os indivíduos foram genotipados e fenotipados – e caso

essa predição seja acurada, o modelo pode ser aplicado em populações de melhoramento relacionadas com a PT (Resende Junior et al., 2012; Almeida Filho et al., 2016).

Um dos fatores importantes no ganho genético é a acurácia da predição; no caso da GS essa acurácia pode ser maximizada usando o modelo que melhor se adira à complexidade genética das características de interesse (Rabier et al., 2016). Uma gama de métodos estatísticos com diferentes pressuposições sobre a arquitetura genética das características está disponível na literatura (Meuwissen et al., 2001; Gianola e Kaam, 2008; De Los Campos, 2009; Legarra et al., 2015). Encontram-se também diversos trabalhos publicados comparando os principais métodos de GS (Pszczola et al., 2011; Heslot et al., 2012; Thavamanikumar et al., 2015; Wang et al., 2015); contudo, ainda não há trabalhos com GS em populações reais de melhoramento com milho-pipoca em que tenham sido realizados estudos comparativos com esses métodos.

Segundo Crossa et al. (2017), a aplicação da GS no melhoramento de plantas pode ser limitada por dois fatores principais: a) custos de genotipagem; e b) diretrizes pouco claras sobre onde a GS pode ser eficientemente aplicada em um programa de melhoramento. O custo da genotipagem é menor à medida que o painel de SNPs se torna menos denso; dessa forma, usando número reduzido de SNPs é possível genotipar maior número de indivíduos com o mesmo recurso, aumentando, assim, a intensidade de seleção e, conseqüentemente, o ganho genético. Apesar da relevância na determinação da densidade correta de SNPs para aplicação rotineira de GS, esse dimensionamento ainda não foi extensivamente abordado na literatura para milho-pipoca.

Posto isso, o presente estudo teve como objetivo estimar as acurácias seletivas utilizando os seguintes métodos: GBLUP, Bayes A, Bayes B, Bayes Cπ, Bayes Lasso e RKHS; e, após definido o melhor método, testar diferentes densidades de marcadores de SNPs visando a verificar se com menor densidade de marcadores SNPs podem-se encontrar valores satisfatórios de acurácia, no intento de otimizar a aplicação da GS no melhoramento genético da cultura do milho-pipoca.

3.2.2. REVISÃO

3.2.2.1. MÉTODOS ESTATÍSTICOS APLICADOS NA GS

Para ajustar um modelo preditivo utilizando a genômica, é preciso inicialmente um grupo de indivíduos genotipados e fenotipados para as características de interesse. Posteriormente, esse modelo pode ser utilizado para prever o mérito genético de indivíduos que nem sequer foram fenotipados, mas que dispõem de informações de marcadores, sendo a seleção genômica a seleção dos indivíduos com base neste mérito genético predito (Goddard e Hayes, 2009).

A escolha de métodos estatísticos para predição dos efeitos de marcadores pode afetar a acurácia de GEBV, sendo que o melhor método é aquele que reflete melhor a natureza biológica do caráter poligênico em questão, em termos de efeitos gênicos (Resende et al., 2008). Desta forma, a correta modelagem da arquitetura genética das características é fundamental. Para tanto, existem diversos métodos para predição genômica, os quais divergem na suposição sobre o modelo genético associado ao caráter quantitativo e são divididos em paramétricos (Meuwissen et al., 2001) e não-paramétricos (Gianola e Van Kaam, 2008). Estes métodos podem ser classificados em três diferentes grupos: regressão explícita, implícita e com redução dimensional, divergindo em relação ao método de estimação utilizado (Tabela 3). A evolução dos métodos de estimação de componentes de médias apresenta-se detalhados em Resende et al., 2012.

Tabela 3 - Principais métodos estatísticos para predição genômica

Regressão Explícita		Regressão Implícita	Redução Dimensional
Estimação Penalizada	Estimação Bayesiana		
RR-BLUP (Meuwissen et al., 2001).	Bayes A, Bayes B (Meuwissen et al., 2001).	Redes neurais artificiais-RNAs (método semi-paramétrico) (Gianola e De Los Campos, 2009).	Componentes independentes.
LASSO (Tibshirani, 1996).	Bayes C, Bayes Σ , Bayes D, Bayes $D\Sigma$ (Habier et al., 2011).	Kernel não-paramétrica, via modelos aditivos generalizados (Gianola et al., 2006).	Quadrados mínimos parciais.
Rede Elástica (Zou e Hastie, 2005).	Lasso Bayesiano (Park e Casella, 2008; De Los Campos et al., 2009).		Componentes principais (Solberg et al., 2009).

Dentre estes métodos de predição genômica, diversas comparações e estudos têm sido realizados. Meuwissen et al. (2001) verificaram uma pequena superioridade do método Bayes B, em relação a RR-BLUP. Honarvar e Rostami (2013) relataram que o método Bayes Lasso apresentou maior acurácia que o método RR-BLUP. Wang et al. (2015), utilizando uma população simulada, encontraram que os métodos Bayes A, Bayes B e Bayes Σ apresentaram acurácias semelhantes e que todos superaram o método GBLUP na estimativa do valor genético. Em contraponto, Costa (2015), utilizando 37.024 marcadores SNPs em cana-de-açúcar, comparou cinco modelos preditivos: GBLUP, Bayes Lasso, Bayes A, Bayes B e 'Bayesian Ridge Regression' e concluiu que os métodos apresentaram valores de acurácia muito semelhantes. Porém, relatou que ao comparar predição utilizando '*pedigree*' e predição utilizando informação genômica, a informação genômica explicou maior proporção da variância genética e teve acurácia mais elevada. Azevedo et al. (2015), em trabalho utilizando métodos de redução dimensional, avaliando caracteres relacionados a carcaça de suínos, encontraram resultados semelhantes entre o método dos componentes principais e

o dos componentes independentes e determinaram que estes foram os de maior acurácia na predição genômica.

3.2.2.2. USO E CONTROLE DE QUALIDADE DE SNPS

Os principais marcadores utilizados nos estudos de GS são os SNPs (Morais Júnior, 2013). Solberg et al. (2008) relatam que o uso de SNPs requer quatro a cinco vezes maior densidade de marcadores do que o uso de microssatélites. Isto se deve à natureza multi-alélica dos microssatélites e bi-alélica dos SNPs. Porém, segundo Caetano et al., (2009), com os avanços da tecnologia a genotipagem de dezenas a centenas de marcadores SNPs, em números elevados de amostras, tem sido realizada com altíssima velocidade e automação. O desenvolvimento de plataformas de alta densidade e em larga escala para a genotipagem via SNPs (Hardenbol et al., 2005) abriu diversas possibilidades na GS. Esta mudança tem contribuído para o aumento da acurácia (Zhao et al., 2015), da mesma forma como o aumento do número de fenótipos, que também eleva as estimativas de acurácia dos efeitos de marcadores (Hayes et al., 2009).

Após a identificação dos SNPs, há a necessidade da validação dos dados e estudos de associação para se avaliar o potencial informativo de cada polimorfismo. Avaliar a qualidade de marcadores e amostras por meio de alguns critérios de controle de qualidade é um passo importante para garantir que as análises estatísticas sejam mais precisas (Yang et al., 2011). Este controle de qualidade ajuda a evitar erros de genotipagem (que podem estar associados à degradação do DNA, contaminação com proteínas e RNA), amostras duplicadas e outros fatores que possam afetar os resultados da análise (Panoutsopoulou et al., 2009). De acordo com Anderson et al. (2010), para filtrar os SNPs mais informativos, são indicados o uso do '*Call rate*' e MAF (frequência do alelo menos comum).

O '*Call Rate*' é a proporção de genótipos por marcador sem dados perdidos. Quando uma grande proporção de SNPs falha em uma amostra significa que há baixa qualidade do DNA, e esta deve ser eliminada da análise (Anderson et al., 2010). Turner et al. (2011) assumem que SNPs que expressarem uma proporção acima de 5 % de falha de genotipagem devem ser eliminados a fim de evitar erros de associação. Em uma análise de associação genômica, alelos com frequência baixa podem conduzir a erros, pois não são informativos. O valor considerado para

remoção dos SNPs pode variar entre 1-2 %; entretanto, em estudos com poucas amostras podem ser necessário valores maiores (Anderson et al., 2010).

Populações sob seleção recorrente estão supostamente em EHW (Equilíbrio de Hardy-Weinberg), devido à recombinação; portanto, pode ser especulada a 'filtragem' de marcadores pelo EHW. Os SNPs que apresentam desvios do EHW podem sinalizar problemas, como erros na determinação do perfil de SNPs ou peculiaridades nos dados, o que tem levado o EHW a ser utilizado no controle de qualidade dos genótipos (Laurie et al., 2010). Em suma, o controle de qualidade é fundamentado na eliminação de marcadores com baixa eficiência de genotipagem, baixa frequência do menor alelo e SNPs que desviam significativamente do EHW (Anderson et al., 2010).

3.2.2.3. DESEQUILÍBRIO DE LIGAÇÃO E ESTRUTURA POPULACIONAL

Para o sucesso da GS é importante conhecer a extensão do desequilíbrio de ligação (LD) ao longo do genoma. Desequilíbrio de ligação (LD) é a associação não aleatória entre alelos de diferentes locos em uma população, sendo a medida r^2 uma das mais utilizadas (Mangin et al., 2011). Idealmente, para a seleção genômica, todos os genes que contribuem para a variação do caráter estarão em desequilíbrio de ligação com pelo menos um marcador, na população em estudo (Goddard e Hayes, 2007). Dessa forma, o decaimento do desequilíbrio de ligação (LD) sobre a distância física em uma população determina a densidade da cobertura por marcadores necessária para realizar análises genômicas. Por exemplo, se o LD decai rapidamente, uma densidade mais elevada de marcadores é necessária para obter marcadores localizados perto o suficiente de locos funcionais (Yu e Buckler, 2006).

Outro fator importante para a GS é a estrutura populacional, pois o efeito da estrutura poderá ter grande impacto na extrapolação do modelo preditivo para outra população não relacionada (Wray et al., 2013). Refere-se à estrutura populacional que ocorre naturalmente devido a processos dispersivos (deriva genética), sistemáticos (mutação, migração e seleção), ou em acasalamentos controlados, comumente observados em populações sob melhoramento.

3.2.3. MATERIAL E MÉTODOS

3.2.3.1. POPULAÇÃO DE ESTUDO

A população em uso teve início com a população UNB-2U, sendo esta uma variedade de polinização aberta. Antes de chegar à UENF (Universidade Estadual do Norte Fluminense Darcy Ribeiro), esta se constituiu de um composto indígena doado à Universidade de Brasília (UNB) pela ESALQ/USP, que recebeu o nome de UNB-1. Posteriormente, a UNB-2 foi trazida para a UENF pelo Professor Joachim Friedrich Wilhelm Von Bülow no ano de 1993, a qual foi cruzada com a variedade de milho-pipoca SAM ('South American Mushroom'). Esta primeira geração filial foi, então, cruzada com uma variedade de milho-pipoca resistente a *Exserohilum turcicum* (helmintosporiose). Após dois ciclos de seleção massal, foram feitos três retrocruzamentos com a variedade SAM, originando, assim, a população UNB-2. Esta, após dois ciclos de seleção massal, deu origem à população UNB-2U (Pereira e Amaral Júnior, 2001). Foram, então, realizados nove ciclos de seleção recorrente visando a obter ganhos para as principais características de importância econômica da cultura – capacidade de expansão e rendimento de grãos. O primeiro ciclo foi constituído por famílias de irmãos-completos (Daros et al., 2002), o segundo ciclo por famílias S₁ (Daros et al., 2004), o terceiro ciclo por famílias de meios-irmãos (Santos et al., 2008) e os demais ciclos por famílias de irmãos-completos (Freitas Júnior et al., 2009; Rangel et al., 2011; Ribeiro et al., 2012; Freitas et al., 2014; Guimarães et al., 2016).

O presente estudo com a população UENF-14 foi obtido após o oitavo ciclo de Seleção Recorrente (SR). Para tanto, após a recombinação das progênies selecionadas, foram semeadas as sementes da população resultante, iniciando o nono ciclo de SR, o qual foi utilizado neste trabalho.

3.2.3.2. GENOTIPAGEM

A partir de 200 amostras de DNA de plântulas coletadas, foi realizada a caracterização do polimorfismo no genoma entre indivíduos dessa população, utilizando a técnica 'Capture Seq', considerando 5.000 sondas bem distribuídas no genoma de referência do milho. Essa abordagem genômica foi realizada em

colaboração com a empresa 'Rapid Genomics' LLC na Florida-USA, resultando em 21.442 SNPs.

De posse das informações genotípicas, com o 'software' Plink (Purcell et al., 2007), foram realizados três filtros, na sequência: a) retirada de indivíduos com > 10 % de dados perdidos; b) descarte de SNPs com > 5 % de dados perdidos e c) descarte de SNPs com MAF (frequência do alelo menos comum) < 5 %. A aplicação desses filtros resultou em 196 indivíduos e 10.507 SNPs; dentre esses, 98 indivíduos foram utilizados no ensaio de progênies S₁.

3.2.3.3. FENOTIPAGEM

Com as famílias S₁ obtidas, foram delineados dois ensaios, sendo um no Colégio Estadual Agrícola Antônio Sarlo, em Campos dos Goytacazes, região Norte do Estado do Rio de Janeiro, situado a 21° 45' de latitude sul e 41° 20' W de longitude e a 11 m de altitude, classificado como tropical chuvoso, clima de bosque com uma precipitação média anual de 1.023 mm, evapotranspiração potencial de 1.601 mm anuais e temperatura média anual de 23° C, e o outro experimento na Estação Experimental da PESAGRO-RIO de Itaocara, localizada na Região Noroeste Fluminense, situada a 21° 39' 12" de latitude sul e 42° 04' 36" W de longitude e a 60 m de altitude, com temperatura média anual de 22,5° C e precipitação média anual de 1.041 mm, ambos implantados em agosto de 2016. O delineamento experimental utilizado foi blocos incompletos, com três repetições. Foi utilizada densidade populacional de 60.000 plantas por hectare. Os tratos culturais foram realizados conforme as necessidades da cultura.

As características fenotipadas, foram: a) rendimento de grãos, expresso em kg.ha⁻¹; e b) capacidade de expansão, determinada em vasilha plástica sem óleo, com três repetições por parcela com peso de 30 g de grãos cada. O pipocamento foi realizado em microondas durante 2 minutos, após o qual, o volume da pipoca expandida foi mensurado em proveta de 2.000 mL e o resultado dividido pela massa inicial de 30 g de grãos e expresso pela relação mL g⁻¹; e c) volume de pipoca expandida por hectare (VP), obtido pela multiplicação entre a produtividade média da parcela e a capacidade de expansão, gerando o volume médio de pipoca expandida por hectare de plantio, expresso em m³.ha⁻¹ (VP= RG x CE/1000). A partir

das observações fenotípicas, foram estimadas as médias ajustadas (LS-Means), corrigidas para umidade de colheita, estande inicial e delineamento experimental.

3.2.3.4. MÉTODOS ESTATÍSTICOS E VALIDAÇÃO CRUZADA

Neste estudo, foram comparados os valores de acurácia de predição para as características em estudo nos seguintes métodos: GBLUP, Bayes A, Bayes B, Bayes Cπ, Bayes Lasso e RKHS. Para observar diferenças entre os métodos foi utilizado o teste t, com intervalo de confiança de 95 %. A validação cruzada 'ten-fold' foi utilizada para verificar a eficácia da seleção (Resende Junior et al., 2012; Almeida Filho et al., 2016). Neste processo, as estimativas são ajustadas com 90 % dos indivíduos da população, enquanto a parte remanescente (10 %) tem o mérito genético predito considerando apenas as informações de marcadores. Esse processo de ajuste de modelo foi repetido dez vezes, sendo que em cada ciclo um grupo diferente de indivíduos teve o mérito genético predito com o fenótipo ignorado.

3.2.3.4.1. GBLUP

O procedimento designado de 'Genomic Best Linear Unbiased Predictor' (GBLUP) é um modelo de indivíduos, onde foi realizada a substituição da matriz de parentesco baseada no '*pedigree*' pela matriz de parentesco estimada por marcadores. Como para a obtenção da GRM (matriz de parentesco genético), todo marcador com mesma MAF contribui igualmente para a variância genética, pode-se dizer que esse modelo é condizente com o modelo de herança infinitesimal, tradicionalmente assumido na genética quantitativa clássica, que pressupõe que as características são poligênicas com número de genes que tende ao infinito e todos contribuem igualmente com a variação fenotípica.

O modelo GBLUP ajustado foi:

$$y = \mu + g + e$$

Em que: y é o vetor de informações fenotípicas (média ajustada) de uma determinada característica, μ é o intercepto do modelo, g é o efeito aleatório do valor genético aditivo genômico (GEBV); e ϵ é o efeito do erro aleatório.

Para este modelo foi assumido que:

$$g \sim N(0, G\sigma_A^2); \epsilon \sim N(0, I\sigma_\epsilon^2),$$

Em que: **I** e **G** são matrizes identidade e GRM, respectivamente. Os componentes de variância foram estimados via REML com o algoritmo AI utilizando o "pacote" asreml (Butler, 2007) do 'software' R (R Core Team, 2013).

3.2.3.4.2. MÉTODOS BAYESIANOS

Alternativamente, foram comparados modelos de regressão que ajustam os SNPs como co-variáveis (De los Campos et al., 2013; Gianola, 2013). A versatilidade dos modelos de regressão advém das diferentes pressuposições assumidas pelos efeitos dos marcadores, que são integradas ao modelo via inferência Bayesiana. Para tanto, assume-se o seguinte modelo geral:

$$y_j = \mu + \sum_{i=1}^{n^o \text{ SNPs}} x_{ij} a_i + \varepsilon_j,$$

Em que: y_j é o valor fenotípico corrigido previamente pelo delineamento experimental (média ajustada); μ é o intercepto; a_i é efeito de substituição alélica do marcador i ; x_{ij} é o número de cópias do alelo menos frequente do indivíduo j para o marcador i (2, 1 ou 0); e ε_j é o efeito do erro aleatório.

De forma geral, pode ser assumido sob a abordagem Bayesiana (Pérez e De los Campos, 2014), o seguinte:

$$\mu \sim N(0, 10^6)$$

$$\varepsilon_j | \sigma_\varepsilon^2 \sim N(0, \sigma_\varepsilon^2)$$

$$\sigma_\varepsilon^2 | gl, S_\varepsilon \sim \chi^{-2}(gl, S_\varepsilon)$$

Para o efeito de substituição alélica (a_i), as pressuposições foram formuladas de acordo com as '*prioris*' assumidas, dentre as diversas '*prioris*' disponíveis na literatura (Habier et al., 2011; De los Campos et al., 2013; Gianola, 2013; Pérez e De los Campos, 2014). Na sequência, segue um breve resumo de cada método testado.

3.2.3.4.3. BAYES A

As pressuposições do Bayes A permitem que marcadores com mesma MAF contribuam de forma diferente para a variância genética, uma vez que as variâncias do efeito das marcas são heterogêneas (Pérez e De los Campos, 2014). O procedimento Bayes A foi proposto por Meuwissen et al. (2001) e assume que:

$$a_i | \sigma_{a_i}^2 \sim N(0, \sigma_{a_i}^2)$$

$$\sigma_{a_i}^2 | gl, S_a \sim \chi^{-2}(gl, S_a)$$

$$S_a | r, s \sim G(r, s)$$

3.2.3.4.4. BAYES B

O método Bayes B pode ser compreendido como um complemento ao Bayes A, visto que além de ajustar os marcadores com variâncias heterogêneas, o Bayes B também assume que algumas marcas não estão em LD com qualquer gene. Assim, têm o efeito zerado. O procedimento Bayes B é formulado por meio de uma mistura de distribuições (Pérez e De los Campos, 2014), considerando-se as seguintes pressuposições:

$$a_i | \sigma_{a_i}^2 \sim N(0, \sigma_{a_i}^2) = 0 \quad \begin{array}{l} \text{com probabilidade } 1-\pi \\ \text{com probabilidade } \pi \end{array}$$

$$\pi | \pi_0, p \sim \text{beta}(\pi_0, p)$$

$$\sigma_{a_i}^2 | gl, S_a \sim \chi^{-2}(gl, S_a)$$

$$S_a | r, s \sim G(r, s)$$

3.2.3.4.5. BAYES LASSO (BL)

De forma similar à concepção dos métodos Bayesianos anteriores, o BL assume variâncias heterogêneas para o efeito das marcas, e também prevê que várias marcas podem não estar em desequilíbrio de ligação com qualquer gene; porém, a seleção de marcas do BL é realizada de forma indireta, por meio da distribuição marginal do efeito de marcas, que é a exponencial dupla (DE) (Park e Casella, 2008; De los Campos et al., 2009). Essa distribuição é mais leptocúrtica do que a distribuição a '*priori*' marginal \mathbf{t} , de Student, usada em Bayes A e B (Gianola et al., 2009). O BL que foi ajustado neste estudo, assume:

$$a_i | \sigma_\varepsilon^2, \tau_i^2 \sim N(0, \sigma_\varepsilon^2 \tau_i^2)$$

$$\tau_i^2 | \lambda \sim \text{Exponencial}(0.5\lambda^2)$$

$$\lambda | r, s \sim \text{Gamma}(r, s)$$

De acordo com Park e Casella (2008) e De los Campos et al. (2009):

$$a_i | \lambda \sim DE(\lambda)$$

3.2.3.4.6. BAYES C π

A metodologia Bayes C π foi proposta por Habier et al. (2011). Nesse método, similar ao Bayes B, é assumido que alguns marcadores não estão em desequilíbrio de ligação com qualquer gene, portanto tem efeito nulo; contudo, entre os SNPs de efeito não nulo, todos apresentam variância homogênea, dessa forma, neste modelo, os SNPs de mesma MAF que estão em desequilíbrio de ligação com genes que governam a característica explicam a mesma proporção da variância genética.

$$\begin{aligned} a_i | \sigma_a^2 &\sim N(0, \sigma_a^2) \quad \text{com probabilidade } 1-\pi \\ &= 0 \quad \text{com probabilidade } \pi \\ \pi | \pi_0, p &\sim \text{beta}(\pi_0, p) \\ \sigma_a^2 | gl, S_a &\sim \chi^{-2}(gl, S_a) \end{aligned}$$

3.2.3.4.7. RKHS

O método de regressão semi-paramétrica 'Reproducing Kernel Hilbert Spaces' (RKHS) é um procedimento que prediz o mérito genético dos indivíduos diretamente, isto é, prescindindo da estimação do efeito de marcas. De fato, o GBLUP é um caso especial de RKHS (De los Campos, 2009); contudo, neste trabalho, o modelo chamado de RKHS foi ajustado seguindo a abordagem 'Kernel Averaging' (De los Campos et al., 2010), usando três funções das informações genotípicas:

$$y = \mu + \sum_{i=1}^3 g_i + e$$

Onde:

$$\begin{aligned} [g_1 \ g_2 \ g_3]' &\sim N(0, \bigoplus_{i=1}^3 K_i \sigma_{g_i}^2) \\ \sigma_{g_i}^2 | \nu, S &\sim \chi^{-2}(\nu, S) \\ K_i &= \exp(-\varphi_i D^2) \end{aligned}$$

\mathbf{D} é a matriz de distância euclidiana considerando as informações de SNPs (número de dose de um determinado alelo); φ_i é o i -ésimo valor ‘bandwidth’ que controla a magnitude da covariância entre os indivíduos, sendo que os valores bandwidth usados foram $5/h$, $1/h$ e $0,2/h$, em que h corresponde ao percentil 0,05 de \mathbf{D} , expandindo para local, intermediária e global gaussian Kernels (González-Camacho et al., 2012; Tusell et al., 2014); \oplus é o operador de soma direta; os outros componentes do modelo já foram explicados anteriormente. Neste modelo, o valor genotípico total (efeitos aditivos e não aditivos) é estimado de forma confundida (Morota e Gianola, 2014) pelo componente $\sum_{i=1}^3 g_i$, não sendo possível realizar o desdobramento ortogonal.

3.2.3.5. **CORRELATION NETWORK’ E TEMPOS DE PROCESSAMENTO COMPUTACIONAL**

Após o ajuste dos modelos de GS foram realizadas análises de correlações em rede (*‘correlation network’*) entre os diferentes métodos, de forma a identificar o método mais adequado. Adicionalmente, foram computados os tempos de processamento computacional de cada método, a fim de comparar o GBLUP com os métodos Bayesianos de maior destaque nos resultados. O modelo frequentista foi ajustado com o pacote ‘rrBLUP’ e os Bayesianos com o pacote ‘BGLR’, do ‘software’ R. Cada método foi analisado dez vezes em ordens aleatórias.

3.2.3.6. **DIFERENTES DENSIDADES DE MARCADORES SNPS**

Depois de identificado o melhor método para a seleção recorrente genômica, foram avaliadas diferentes densidades de marcadores SNPs, utilizando-se o desequilíbrio de ligação. O desequilíbrio de ligação (LD) é definido como a associação não-aleatória entre alelos de diferentes locos gênicos, e é calculado entre a diferença das frequências observadas e esperadas dos haplótipos, considerando a segregação independente dos alelos (Weiss e Clark, 2002).

Para determinar a possibilidade de usar painéis menos densos de SNPs, foram realizados filtros por meio da abordagem LD *‘pruning’*, que consiste em manter apenas um dos SNPs, caso dois SNPs tenham LD acima de um ponto de

corte; dessa forma, foram adotados pontos de corte para estatística de LD r^2 de 0,05 (resultando em 418 SNPs) até 1 (painel completo com 10.507 SNPs), considerando toda extensão do cromossomo nas comparações e filtros.

Obtiveram-se, então, as acurácias seletivas para as diferentes características em estudo, de forma a detectar a menor densidade de SNPs com acurácia satisfatória. O LD '*prunning*' foi realizado com o 'software' Plink (Purcell et al., 2007).

3.2.4. RESULTADOS E DISCUSSÃO

3.2.4.1. COMPARAÇÃO DE MÉTODOS

As acurácias seletivas dos diferentes métodos testados foram iguais estatisticamente pelo teste t, com intervalo de confiança de 95 %. Para a característica RG, as acurácias se situaram entre 0,1922 para Bayes Lasso e 0,2685 para Bayes A, com média geral de 0,2428. Pode-se observar que os métodos Bayes A, Bayes B, Bayes C π e GBLUP proporcionaram resultados semelhantes, com estimativas de acurácia em torno de 0,26. Os métodos BLasso e RKHS apresentaram resultados inferiores, proporcionando estimativas de acurácias que se situaram entre 0,1922 e 0,2030, respectivamente (Figura 4). Wang et al. (2015) também encontraram resultados semelhantes com modelos diferentes para RG em trigo. Riedelsheimer et al. (2012) não encontraram grandes diferenças entre Bayes B e outros modelos na predição de várias características, incluindo caracteres com QTLs de grandes efeitos em milho.

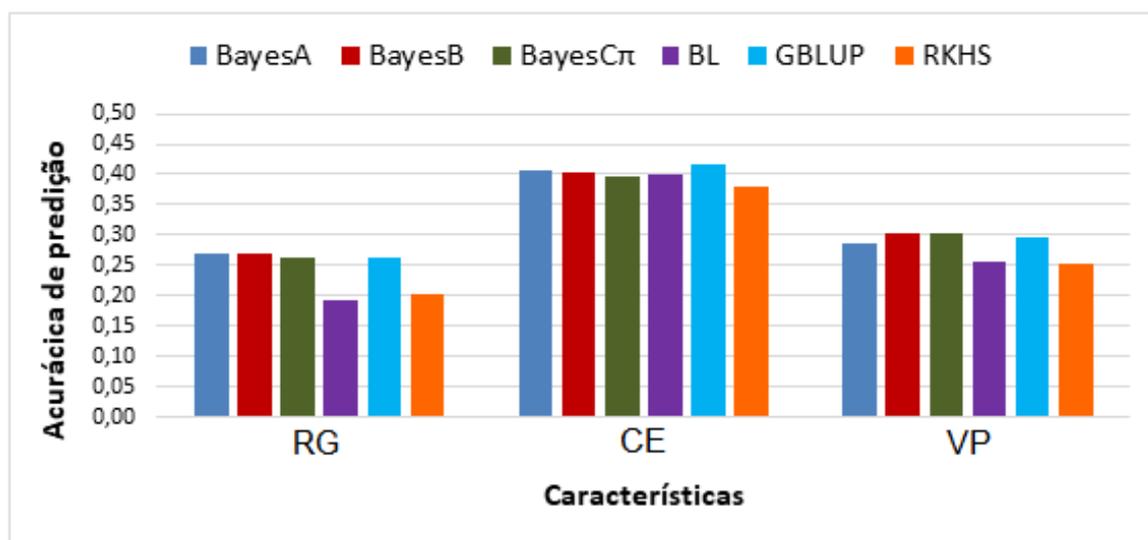


Figura 4 - Estimativas das acurácias preditivas das médias obtidas para os métodos Bayes A, Bayes B, Bayes C, Bayes Lasso, GBLUP e RKHS. Estimativas das acurácias preditivas das médias obtidas para os métodos Bayes A, Bayes B, Bayes Cπ, Bayes Lasso, GBLUP e RKHS em população de milho-pipoca para as características rendimento de grãos (RG), capacidade de expansão (CE) e volume de pipoca (VP).

Para CE, as acurácias se situaram entre 0,3811 para RKHS e 0,4164 para o método frequentista GBLUP, com estimativa de média geral de 0,4007. Os métodos Bayes Cπ, Bayes Lasso e RKHS apresentaram resultados numericamente inferiores, com estimativas próximas de 0,39. GBLUP se destacou, com acurácia de 0,4164 (Figura 4). Xu et al. (2018) apontaram que o GBLUP foi superior aos métodos Bayesianos na predição de seis características de milho. Na expressão gênica de CE tem sido reportada por alguns autores a predominância de efeitos aditivos (Larish e Brewbaker, 1999; Pereira e Amaral Junior, 2001; Freitas Júnior et al., 2006; Scapim et al., 2006; Rangel et al., 2008). Segundo Wang et al. (2015), os métodos de GS existentes baseiam-se principalmente em modelo aditivo e ainda é difícil estimar com precisão a variação não-aditiva. Isto posto, pode-se inferir que os melhores valores de acurácia encontrados para CE sejam explicados pela maior herdabilidade, ou seja, menor influência ambiental na manifestação do caráter.

Já para VP, as acurácias revelaram valores entre 0,2517 para a RKHS e 0,3037 para Bayes Cπ, com média geral de 0,2829. Os métodos Bayes B, Bayes Cπ e GBLUP proporcionaram resultados semelhantes, com valores estimados próximos a 0,30. Os métodos Bayes Lasso e RKHS, por sua vez, revelaram estimativas inferiores, com valores próximos a 0,25 (Figura 4), sendo, portanto,

menos acurados. Segundo Amaral Junior et al. (2016) a utilização de VP como um super-caráter contorna o desafio enfrentado pelos melhoristas de milho-pipoca quanto à ocorrência de correlação negativa entre CE e RG, pois VP permite obter ganhos simultâneos para as duas principais características de importância econômica para a cultura. A característica VP expressou estimativas de acurácia intermediárias em relação a CE e RG, e assim como ocorreu para com RG e CE, os métodos GBLUP, Bayes A e Bayes B apresentaram os resultados mais satisfatórios para fins seletivos.

Nas '*correlation networks*', as variáveis são representadas por pontos que estão conectados por setas, em que cada seta possui diferente espessura, a qual está diretamente vinculada com a intensidade da correlação. O quão mais forte for à correlação entre duas variáveis, mais espessa será a linha conectando os pontos na trama da rede (Silva et al., 2016). As '*correlation networks*' têm sido utilizadas para caracterizar sistemas complexos em diversas áreas, como biologia (DiLeo et al., 2011; Pearce et al., 2015), saúde pública (Saba et al., 2014), ciência de alimentos (Monforte et al., 2015) e melhoramento (Silva et al., 2016). No presente estudo, a correlação entre os resultados obtidos para as diferentes características nos diferentes métodos foi superior a 0,98 (Figura 5). Esse resultado é mais um indício de que não existem diferenças significativas entre os diferentes métodos testados.

Heslot et al. (2012), em estudo comparativo entre diferentes métodos, analisaram o tempo de computação para realizar estudos de seleção genômica. No presente estudo, analisando-se o tempo de computação para obter os resultados por cada método pode-se verificar que foi consideravelmente menor para o GBLUP, apresentando resultados quase que instantâneos. Os modelos Bayesianos demoraram cerca de quatro minutos, enquanto o GBLUP pôde ser obtido em menos de um segundo (Tabela 4). Segundo Wang et al. (2018), a estimativa de métodos Bayesianos é geralmente demorada, restringindo, pois, sua aplicação.

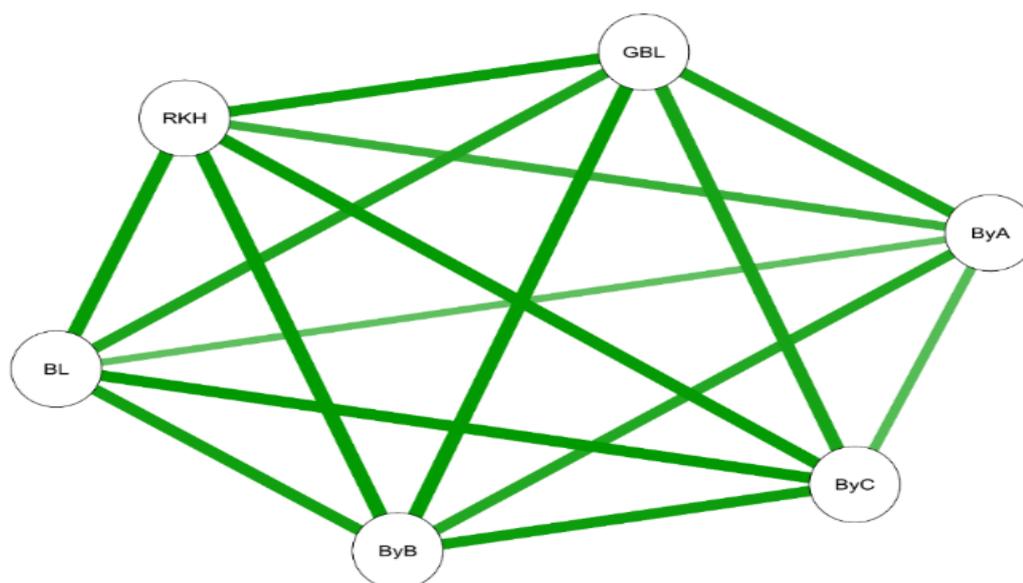


Figura 5 - 'Correlation networks' obtidas nos métodos GBLUP, Bayes A, Bayes B, Bayes π , Bayes Lasso e RKHS. 'Correlation networks' obtidas nos métodos GBLUP, Bayes A, Bayes B, Bayes π , Bayes Lasso e RKHS, para as características rendimento de grãos (RG), capacidade de expansão (CE) e volume de pipoca (VP). Significativo em $p < 0,05$.

Em suma, considerando os resultados encontrados pelos diferentes métodos para as características em estudo, pode-se concluir que os métodos não expressaram diferenças estatísticas significativas. No entanto, os métodos GBLUP, Bayes A e Bayes B destacaram-se com estimativas de acurácias superiores numericamente para todas as características em estudo. Considerando este fato e que o método GBLUP apresentou vantagem em relação à simplicidade de execução com menor tempo demandado na análise, além de proporcionar maior acurácia numérica para a principal característica de qualidade da pipoca – CE –, pode-se indicar este método para uso na Seleção Recorrente Genômica com a cultura do milho-pipoca.

Tabela 4 - Estimativa de tempo gasto para obtenção de resultados pelos métodos Bayes A, Bayes B e GBLUP.

Modelo	Tempo (segundos)	
	Média	Desvio-padrão
Bayes A	229,3784	1,2024
Bayes B	282,6650	1,7503
GBLUP	0,0143	0,0038

3.2.4.2. DIFERENTES DENSIDADES DE SNPs

Na seleção genômica a estimação do desequilíbrio de ligação deve ser uma das primeiras análises a serem realizadas, a fim de investigar a informatividade dos marcadores utilizados. De acordo com Resende et al. (2008) a variância genética de uma característica de caráter quantitativo pode ser explicada pela presença dos marcadores em desequilíbrio de ligação com os QTLs – sejam esses de pequenos, ou de grandes efeitos – uma vez que somente os marcadores em desequilíbrio são utilizados na determinação dos fenótipos via genótipo.

Considerando-se que as premissas do GBLUP proporcionaram os resultados mais satisfatórios, procedeu-se, então, a avaliações de diferentes densidades de SNPs, no intento de identificar a de maior benefício, tendo-se em conta o custo considerável para a geração de maior quantidade de marcas. Assim, foram testadas 20 diferentes densidades de SNPs – variando de 10.507 ($r^2 = 1,00$) a 418 ($r^2 = 0,05$) –, utilizando-se o LD (r^2) como filtro (Figura 6). Por exemplo, para 6.766 SNPs o ponto de corte foi $r^2 = 0,70$, ou seja, se dois marcadores apresentaram $r^2 > 0,70$ apenas um foi mantido.

Para a característica RG, o maior valor de acurácia (0,303) foi encontrado com uso de 6.766 SNPs ($r^2 = 0,70$); no entanto, com 4.848 SNPs ($r^2 = 0,45$) obteve-se acurácia semelhante (0,296). Para PV, o cenário com 4.848 SNPs ($r^2 = 0,45$) também apresentou resultado expressivo de acurácia, com estimativa de 0,311 (Figura 6). Para RG e VP, quando se utilizam densidades inferiores a 4.406 SNPs ($r^2 = 0,40$) ocorre decréscimo nas acurácias, ao contrário do visualizado com a acurácia para densidade de 418 SNPs ($r^2 = 0,05$). Todavia, a elevação da acurácia com densidade de 418 SNPs ($r^2 = 0,05$) certamente se trata de um resultado falso-positivo. Nesse aspecto, Dickson et al. (2011) afirmaram que variantes raras podem

dar origem a taxas inaceitáveis de falso-positivos. Segundo Litonjua e Celedo (2006), são três as possíveis condicionantes para que uma associação entre um polimorfismo e um fenótipo seja significativa: a) ocorrência de associação espúria; b) proximidade de uma variante funcional em um loco adjacente (LD); e c) o loco é funcional e afeta diretamente a expressão do fenótipo. Segundo Utsunomiya et al. (2013), pequenos erros de determinação dos genótipos também podem resultar em níveis inaceitáveis de erros do tipo I (falso-positivo) e II (falso-negativo).

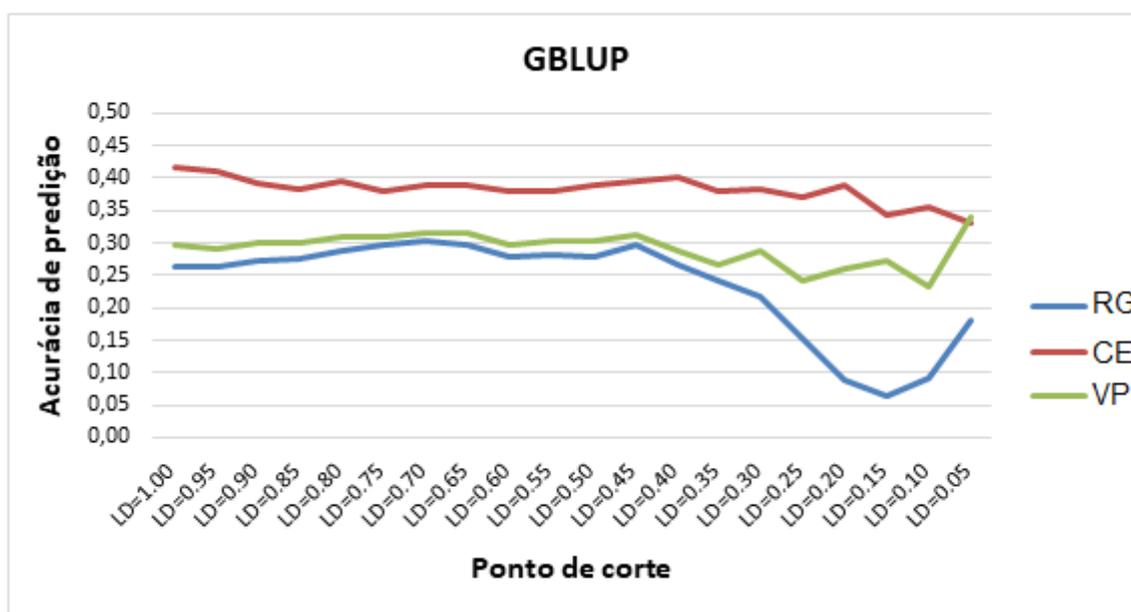


Figura 6 - Gráficos que expressam as diferentes acurácias. Gráficos que expressam as diferentes acurácias de predição obtidas para diferentes densidades de SNPs pelo método GBLUP para as características rendimento de grãos (RG), capacidade de expansão (CE) e volume de pipoca (VP).

Para CE, a maior acurácia (0,416) foi encontrada com 10.507 SNPs ($r^2 = 1,00$); porém, com uso de 4.406 SNPs ($r^2 = 0,40$) obteve-se acurácia semelhante (0,402). Para CE a acurácia apresentou variação de 0,416 (10.507 SNPs) a 0,332 (418 SNPs), sendo esta bem inferior à variação ocorrida para a característica RG (Figura 6). Wang et al. (2015), trabalhando com dados de trigo do 'Wheat Global Program' do Centro Internacional de Melhoramento de Milho e Trigo (CIMMYT), também observaram que a precisão e capacidade preditiva do GBLUP

permaneceram quase constantes para caracteres aditivos, independentemente do número de QTLs utilizados.

Segundo Hiremath et al. (2012) os marcadores SNPs são receptivos à automação e à abordagem de alto rendimento e os custos de genotipagem podem ser reduzidos. Considerando os resultados encontrados nos diferentes cenários de densidade de SNPs avaliados no presente trabalho, pode-se inferir que com o uso de ~4.800 SNPs ($r^2 = 0,45$), podem-se obter resultados semelhantes ou até melhores que os encontrados com 10.507 SNPs ($r^2 = 1,00$). Posto isso, torna-se possível uma redução de aproximadamente 50 % na densidade de SNPs utilizada, o que resulta em significativa redução de custos na seleção recorrente genômica em milho-pipoca.

3.2.5. CONCLUSÕES

O método GBLUP pode ser recomendado com sucesso. Considerável redução de custos pode ser obtida usando um painel reduzido de SNPs.

REFERÊNCIAS BIBLIOGRÁFICAS

- AGUIAR C, Schuster, I, Amaral Junior, A.T., Scapim, C.A., Vieira, E.S.N. (2008) Heterotic groups in tropical maize germplasm by test crosses and simple sequence repeat markers. *Genetics and Molecular Research*, 7(4): 1233-1244.
- ALBRECHT. T, Auinger H. J, Wimmer V, Ogutu J. O, Knaak C, Ouzunova M, Piepho HP, Schön C. C. (2014) Genome-based prediction of maize hybrid performance across genetic groups, testers, locations, and years. *Theoretical and Applied Genetics*, 127: 1375-1386.
- ALMEIDA Filho J. E, Guimarães J. F. R, Silva F. F, Resende M. D. V, Muñoz P, Kirst M, Resende Jr M. F. R. (2016) The contribution of dominance to phenotype prediction in a pine breeding and simulated population. *Heredity* 117:33-41.
- AMARAL Júnior AT, Freitas Júnior SP, Rangel RM, Pena GF, Ribeiro RM, Morais RC, Schuelter AR. (2010) Improvement of a popcorn population using selection indexes from a fourth cycle of recurrent selection program carried out in two different environments. *Genetics and Molecular Research*, 9(1): 340-347.
- AMARAL Junior AT, Gonçalves LSA, Freitas Júnior SP, Candido LS, Vittorazzi C, Pena, GF, Ribeiro, RM, Silva, TRC, Pereira, MG, Scapim, CA, Viana, AP, Carvalho, GF. (2013) UENF 14: a new popcorn cultivar. *Crop Breeding and Applied Biotechnology*. Viçosa, 13:218.

- AMARAL. Júnior A. T *et al.* (2016) Bayesian analysis of quantitative traits in popcorn (*Zea mays* L.) through four cycles of recurrent selection. *Plant Production Science*; 19: 1-5.
- AMARAL. Júnior A. T, Santos A, Gerhardt I. F. S, Kurosawa R. N, Moreira N. F, Pereira M.G., Gravina G.A., de L Silva F.H. (2016) Proposal of a super trait for the optimum selection of popcorn progenies based on path analysis. *Revista Genetic and Molecular Research*, 14(4): gmr15049309.
- ANDERSON C.A., Pettersson F.H., Clarke G.M., Cardon L.R., Morris P, Zonderva K.T. (2010) Data quality control in genetic casecontrol association studies. *Nature Protocols*, 5: 1564-1573.
- ARNHOLD, E. (2008) Seleção para resistência a doenças foliares em famílias S1 de milho-pipoca. *Revista Ceres*, Universidade Federal de Viçosa - Viçosa, 55(2):89-93.
- AZEVEDO., C.F., Nascimento, M., Silva, F.F; Resende, M.D.V (2015) Comparison of dimensionality reduction methods to predict genomic breeding values for carcass traits in pigs. *Genet. Mol. Res.* 14: 1 2217-12227.
- BERNARDO. R, Yu J. (2007) Prospects for genome wide selection for quantitative traits in maize. *Crop Science*; 47: 1082–1090.
- BORÉM. A, Miranda GV, Fritsche-Neto R. (2017) *Melhoramento de plantas*. 7. ed. Viçosa, UFV, 543p.
- BUTLER. D, Cullis BR, Gilmour A. (2007) Asreml-R: an R package for mixed models using residual maximum likelihood. Accessed on 07 December 2018.
- CABRAL. PDS, Amaral Júnior AT, Freitas ILJ, Ribeiro RM, Silva TRC. (2016) Relação causa e efeito de caracteres quantitativos sobre a capacidade de expansão do grão em milho-pipoca. *Revista Ciência Agronômica*; 47(1):108-117.
- CAETANO, Alexandre Rodrigues. (2009) Marcadores SNP: conceitos básicos, aplicações no manejo e no melhoramento animal e perspectivas para o futuro. *R. Bras. Zootec.* [online]. vol.38, pp.64-71.

- COMBS E., Bernardo R. (2013) Accuracy of genome wide selection for different traits with constant population size, heritability, and numbers of markers. *Plant Genome*; 6: 1–7.
- COSTA., P.M.A. (2015) Prediction breeding values in sugarcane using pedigree and genomic information. Disponível em: <<http://www.locus.ufv.br/handle/123456789/7539> > acesso em 20/12/2018.
- CROSSA J, Campos G, Pérez P, Gianola D, Burgueño J, Arus, JL, Makumbi, D, Singh, RP, Dreisigacker S, Yan J, Arief V, Bazinger M, Braum HJ. (2010) Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. *Genetics*, 186: 713–724.
- CROSSA. J, Perez-Rodriguez P.J., Cuevas O, Montesinos-Lopez D.J., De Los Campos G, Burgueno J, Gonzalez-Camacho J.M., Perez-Elizalde S, Beyene Y, Dreisigacker S, Singh R, Zhang X.C., Gowda M. (2017) Roorkiwal M, Rutkoski J, Varshney RK. Genomic selection in plant breeding: methods, models, and perspectives. *Trends Plant Science*, 22:961–975.
- CRUZ. C. D, Regazzi A.J., Carneiro P.C.S. (2012) Modelos biométricos aplicados ao melhoramento genético. Viçosa; UFV.
- DAETWYLER. H. D, Calus MPL, Pong-wong R, De Los Campos G, HICKEY J.M. (2013) Genomic prediction in animals and plants: simulation of data, validation, reporting, and benchmarking. *Genetics*, 193: 347-365
- DAROS. M, Amaral Júnior A.T., Pereira M.G. (2002) Genetic gain for grain yield and popping expansion in full-sib recurrent selection in popcorn. *Crop Breeding and Applied Biotechnology*, 2 (3): 339-344
- DAROS. M, Amaral Júnior A.T., Pereira M.G., Santos F.S., Gabriel A.P.C., Scapim C.A., Freitas Júnior S.P., Silvério L. (2004) Recurrent selection in inbred popcorn families. *Scientia Agricola*, 61(6):609-614.
- DE LOS CAMPOS G, Gianola D, Rosa G.J.M., Wiegand K.A., Crossa J. (2010) Semi-parametric genomic-enabled prediction of genetic values using reproducing kernel Hilbert spaces methods. *Genetics Research* 92(4): 295-308

- DE LOS CAMPOS G, Hickey J.M., Pong-Wong R, Daetwyler H.D., Calus M.P.L. (2013) Whole-genome regression and prediction methods applied to plant and animal breeding. *Genetics* 193:327-345
- De LOS CAMPOS G, Naya H, Gianola D, Crossa J, Legarra A, Manfredi E, Weigel K, Cotes J.M. (2009) Predicting quantitative traits with regression models for dense molecular markers and pedigree. *Genetics* 182(1):375-385
- DICKSON S.P., Wang K, Krantz I, Hakonarson H, Goldstein D.B. (2011) Rare variants create synthetic genome-wide associations. *PLOS BIOLOGY* 8(1): e1000294
- DILEO M.V., Strahan G.D., den Bakker M, Hoekenga O.A. (2011) Weighted correlation network analysis (WGCNA) applied to the tomato fruit metabolome. *PLOS ONE*, 6(10): e 26683.
- ENDELMAN J.B. (2011) Ridge regression and other kernels for genomic selection with R package rrBLUP. *Plant Genome*, 4:250-255.
- ERBE M, Pimentel E.C.G., Sharifi A.R., Simianer H. (2010) Assessment of cross-validation strategies for genomic prediction in cattle. 9th World Congress of Genetics Applied to Livestock Production: 2009; Leipzig., Gesellschaft für Tierzuchtwissenschaften e. V, Giessen, Germany.
- ERTL, J., A. Legarra, Z.G. Vitezica, L. Varona, C. Edel, et al. (2014) Genomic analysis of dominance effects on milk production and conformation traits in Fleckvieh cattle. *Genet. Sel. Evol.* 46(1): 1–10.
- FALCONER D.D. (1981) Introdução a genética quantitativa. Viçosa, MG: UFV.
- FREITAS I.L.J., Amaral Junior A.T., Freitas Júnior S.P., Cabral P.D.S., Ribeiro R.M., Goncalves, L.S.A. (2014) Genetic gains in the UENF-14 popcorn population with recurrent selection. *Genetics and Molecular Research*, 13: 518-527.
- FREITAS Junior S.P., Amaral Júnior A.T., Pereira M.G., Cruz C.D., Scapim C.A. (2006) Capacidade combinatória em milho pipoca por meio de dialelo circulante. *Pesquisa Agropecuária Brasileira* 41:1599-1607

- FREITAS Júnior S.P., Amaral Júnior A.T., Rangel R.M., Viana A.P. (2009) Genetic gains in popcorn by full-sib recurrent selection. *Crop Breeding and Applied Biotechnology*, 9(1): 1-7.
- FRITSCHÉ Neto R, Resende M.D.V., Miranda G.V., Vale J.C. (2012) Seleção genômica ampla e novos métodos de melhoramento do milho. *Revista Ceres*, 59: 794-802.
- GALVÃO J.C.C., Sawazaki E, Miranda G.V. (2000) Comportamento de híbridos de milho pipoca em Coimbra. *Revista Ceres*, 47 (270): 201-218.
- GIANOLA D. (2013) Priors in whole-genome regression: the Bayesian alphabet returns. *Genetics* 90:525-540.
- GIANOLA D., De los Campos G.; Hill W.G., Manfredi E., Fernando R. (2009) Additive genetic variability and the Bayesian alphabet. *Genetics* 183:347-363.
- GIANOLA D., Fernando R.L., Stella A. (2006) Genomic-assisted prediction of genetic value with semiparametric procedures. *Genetics*, 173(1):1761–1776.
- GIANOLA D., van Kaam J.B.C.H.M. (2008) Reproducing kernel Hilbert spaces regression methods for genomic assisted prediction of quantitative traits. *Genetics* 178:2289-2303.
- GILMOUR, A. R., Gogel B. J., Cullis B. R., Welham S. J., Thompson R. (2015) ASReml User Guide Release 4.1 Structural Specification, VSN International Ltd, Hemel Hempstead, HP1 1ES, UK.
- GODDARD M.E., Hayes B.J. (2009) Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nat Rev Genet* 10:381–391.
- GODDARD ME, Hayes BJ (2007) Genomic selection. *J Anim Breed Genet* 124:323–330.
- GONZÁLEZ-CAMACHO J.M., de los Campos G, Pérez P. (2012) Genome-enabled prediction of genetic values using radial basis function neural networks. *Theoretical and Applied Genetics* 125:759-71

- GUIMARÃES A.G., Amaral Júnior A.T., Lima V.J., Leite J.T., Scapim C.A., Vivas M. (2018) Genetic Gains and Selection Advances of the UENF-14 Popcorn Population. *Revista Caatinga*, 31:271-278.
- GUIMARÃES, A. G. (2016) *Ganhos genéticos por seleção recorrente e estreitamento genético na população UENF-14 de milho-pipoca por marcadores SSR-Ests*. Tese (Doutorado em Genética e Melhoramento de Plantas), Campos dos Goytacazes, Universidade Estadual do Norte Fluminense, 89p.
- HABIER D., Fernando R., Kizilkaya K., Garrick D. (2011) Extension of the Bayesian alphabet for genomic selection. *BMC Bioinformatics* 12(1): 186.
- HALLAUER A.R., Miranda Filho J.B., Carena M.J. (2010) *Quantitative genetics in maize breeding*. 3. ed, New York: Springer, 662p.
- HARDENBOI P, Yu F, Belmont J (2005) Highly multiplexed molecular inversion probe genotyping: over 10,000 targeted SNPs genotyped in a single tube assay. *Genome Res.* 15:269–275.
- HEFFNER E.L., Lorenz A.J., Jannink J.L., Sorrells M.E. (2010) Plant Breeding with Genomic Selection: Gain per Unit Time and Cost. *Crop Science*, 50:1681.
- HESLOT N., Jannink J.-L., Sorrells M.E. (2015) Perspectives for genomic selection applications and research in plants. *Crop Science*, 55:1-12.
- HESLOT N., Yang H.P., Sorrells M.E., Jannink J.L. (2012) Genomic selection in plant breeding: A comparison of models. *Crop Science*. 52:146-160
- HILL W.G., Weir B.S. (1988) Variances and covariances of squared linkage disequilibria in finite populations. *Theor Popul Biol.*, 33:54–78.
- HIREMATH P.J., Kumar A., Penmetsa R.V. (2012) Large-scale development of cost-effective SNP marker assays for diversity assessment and genetic mapping in chickpea and comparative mapping in legumes. *Plant Biotechnology Journal* 10(6):716-32.
- HONARVAR M, Rostami M. (2013) Accuracy of genomic prediction using RR-BLUP and Bayesian LASSO. *Eur J Exp Biol*, 3:42–47.

- JANNINK J.L., Lorenz A.J., Iwata H. (2010) Genomic selection in plant breeding: From theory to practice. *Briefings in Functional Genomics*, 9(2):166–177.
- JONAS E., Koning D.J. (2013) 'Does genomic selection have a future in plant breeding?' *Trends in biotechnology*, 9:497–504.
- KASHIANI P., Saleh G. (2010) Estimation of genetic correlations on sweet corn inbred lines using SAS mixed model. *Am J Agr Biol Sci*, 5:309–314.
- KIST, B. B.; FILTER, F. C.; SANTOS, C. E.; CARVALHO, C. (2018) Anuário brasileiro domilho. *Santa Cruz do Sul: Editora Gazeta Santa Cruz*, 88p.
- LANDE R., Thompson R. (1990) Efficiency of marker assisted selection in the improvement of quantitative traits. *Genetics*, 124:743-756.
- LARISH L.L.B., Brewbaker J.L. (1999) Diallel analyses of temperate and tropical popcorns. *Maydica*, 44(4):279-284.
- LAURIE, C.C. (2010) Quality control and quality assurance in genotypic data for genome- wide association studies. *Genet. Epidemiology*, 34:591–602.
- LEGARRA A., Croiseau P., Sanchez M.P., Teysseire S., Salle G., Allais S., Fritz S., Moreno C.R., Ricard A., Elsen J.M. (2015) A comparison of methods for whole-genome QTL mapping using dense markers in four livestock species. *Genetics Selection Evolution*, 47:6.
- LENZ PRN, et al. (2017) Factors affecting the accuracy of genomic selection for growth and wood quality traits in an advanced-breeding population of black spruce (*Picea mariana*). *BMC Genomics*. 18: 335.
- LIAN L., Jacobson A., Zhong S.Q., Bernardo R. (2014) Genomewide prediction accuracy within 969 maize biparental populations. *Crop Science*. 54, 1514–1522.
- LIMA L.P. (2017) *Seleção genômica não paramétrica via distância genética entre subpopulações*. Dissertação (Mestrado em Estatística Aplicada e Biometria) – Viçosa – MG – Universidade Federal de Viçosa – UFV, 101p.

- LITONJUA A.A., Celedo J.C. (2006) Gene Association Studies. In: Encyclopedia of Respiratory Medicine. Oxford: Academic Press, London, pp 240-247.
- LIU H., Zhou H., Wu Y., Li X., Zhao J., Zuo T. et al. (2015) The impact of genetic relationship and linkage disequilibrium on genomic selection. *PLOS ONE*, 10(7): e0132379.
- LORENZ A.J., Smith K.P., Jannink J.L. (2012) Potential and Optimization of Genomic Selection for Fusarium Head Blight Resistance in Six-Row Barley. *Crop Science*, 52:1609–1621.
- MAFRA G.S., Amaral Júnior A.T., Almeida Filho J.E., Vivas M., Santos P.H.A.D., et al. (2019) SNP-based mixed model association of growth- and yield-related traits in popcorn. *PLOS ONE*, 14(6): e0218552.
- MANGIN B, Siberchicot A, Nicolas S, Doligez A, This P, Cierco-Ayrolles C (2011) Novel measures of linkage disequilibrium that correct the bias due to population structure and relatedness. *Heredity*, New York, 108(3):285-291.
- MAPA – Ministério da Agricultura e Pecuária (2019) CULTIVAR WEB http://sistemas.agricultura.gov.br/snpc/cultivarweb/cultivares_registradas.php. Accessed 07 June 2019.
- MATTA F.P., Viana J.M.S. (2003) Eficiências relativas dos processos de seleção entre e dentro de famílias de meios-irmãos em população de milho pipoca. *Ciência e Agrotecnologia, Lavras*, 27 (3): 548-556.
- MENDES de Paula T.O., Amaral Junior A.T., Gonçalves L.S.A., Scapim C.A., Peternelli L.A., Silva V.Q.R.S. (2010) Pi statistics underlying the evaluation of stability, adaptability and relation between the genetic structure and homeostasis in popcorn. *Acta Scientiarum Agronomy*, 32(2): 269-277.
- MEUWISSEN T., Hayes B., Goddard M. (2001) Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. *Genetics*, 157:1819-1829.
- MEUWISSEN T., Luan T., Woolliams J.A. (2011) The unified approach to the use of genomic and pedigree information in genomic evaluations revisited. *Journal Animal Breeding Genetics*, Noruega, 128(6):429-439

- MISZTAL I. (2016) Inexpensive Computation of the Inverse of the Genomic Relationship Matrix in Populations with Small Effective Population Size. *Genetics*, 202(2):401-409.
- MISZTAL I., Legarra A. (2017) Invited review: efficient computation strategies in genomic selection. 11(5): 731-736.
- MONFORTE A.R., Jacobson D., Ferreira A.C.S. (2015) Chemiomics: network reconstruction and kinetics of Port wine aging. *Journal of Agricultural and Food Chemistry* 63(9):2576-2581.
- MORAIS Júnior O.P. (2013) *Seleção Genômica Ampla (GWS) aplicada ao melhoramento populacional*. Escola de Agronomia e Engenharia de Alimentos, Revisão bibliográfica – Goiânia – GO, Universidade Federal de Goiás, 38p.
- MOROTA G., Gianola D. (2014) Kernel-based whole-genome prediction of complex traits: a review. *Frontier Genetics* 5:363.
- MÜLLER D., Schopp P., Melchinger A.E. (2017) Persistency of prediction accuracy and genetic gain in synthetic populations under recurrent genomic selection. *G3*, 7(3):801–811.
- MURANTY H., Troglio M., Sadok I.B., Rifaï M.A., Auwerkerken A., Banchi E. et al. (2015) Accuracy and responses of genomic selection on key traits in apple breeding. *Hortic Res*, 2:15060.
- NORMAN A., Taylor J., Tanaka E., Telfer P., Edwards J., Martinant J.P. et al. (2017) Increased genomic prediction accuracy in wheat breeding using a large Australian panel. *Theor. Appl. Genet*, 130: 2543–2555.
- PANOUSOPOULOU L., Zeggini E. (2009) Finding common susceptibility variants for complex disease: past, present and future. *Briefings in functional Genomics and Proteomics*, 8: 345- 352.
- PARK T., Casella G., (2008) The Bayesian LASSO. *Journal of the American Statistical Association*, 103(482): 681–686.
- PARK T., Cassella G. (2008) The Bayesian Lasso. *Journal of the American Statistical Association*, 103:681-686.

- PEARCE S., Ferguson A., King J., Wilson Z.A. (2015) FlowerNet: a gene expression correlation network for anther and pollen development. *Plant Physiology* 167:1717-1730.
- PEIXOTO, L.A (2013) Redes neurais artificiais na predição do valor genético. Dissertação mestrado em Genética e Melhoramento de Plantas. Universidade Federal de Viçosa.
- PENA G.F., Amaral Júnior A.T.D., Gonçalves L.S.A., Vivas M., Ribeiro R.M., Mafra G.S., Scapim C.A. (2016) Comparison of testers in the selection of S3 families obtained from the UENF-14 variety of popcorn. *Bragantia*, 75(2), 135-144.
- PEREIRA M.G., Amaral Júnior A.T. (2001) Estimation of genetic components in popcorn based on the nested design. *Crop Breeding and Applied Biotechnology*, 1 (1): 3-10.
- PÉREZ P., de los Campos G. (2014) Genome-wide regression and prediction with the BGLR statistical package. *GENETICS*, 198: 483-495.
- PIMENTEL, A.J.B. et al. (2014) Estimação de parâmetros genéticos e predição de valor genético aditivo de trigo utilizando modelos mistos. *Pesq. agropec. Bras*, 49(11): 882-890.
- POLIZEL A.C., Juliatti F.C., Hamawaki O.T., Hamawaki R.L., Guimarães S.L. (2013) Adaptabilidade e estabilidade fenotípica de genótipos de Soja no estado do Mato Grosso. *Bioscience Journal*, 29: 910-920.
- PRITCHARD J.K., Stephens M., Donnelly P. (2000) Inference of Population Structure Using Multilocus Genotype Data. *GENETICS*, 155(2): 945-959
- PSZCZOLA, M, Mulder HA, Calus, MPL (2011) Effect of enlarging the reference population with (un)genotyped animals on the accuracy of genomic selection in dairy cattle. *Journal of Dairy Science*, 94:431–441.
- PURCELL S., Neale B., Todd-Brown K., Thomas L., Ferreira M.A., Bender D., Maller J., Sklar P., de Bakker P.I., Daly M.J., Sham P.C. (2007) PLINK: a toolset for whole-genome association and population-based linkage analysis. *American Journal of Human Genetics*, 81(3):559-75.

- PURCELL, S., and C. Chang. (2015). Plink, Version 1.9.
- R Core Team (2013) R: A language and environment for statistical computing. R Foundation for Statistical Computing. Accessed 07 December 2018.
- RABIER C.E., Barre P., Asp T., Charmet G., Mangin B. (2016) On the Accuracy of Genomic Selection. *PLOS ONE*, 11(6): e0156086.
- RANGEL R.M., Amaral Júnior A.T., Gonçalves L.S.A., Freitas Júnior S.P., Candido L.S. (2011) Análise biométrica de ganhos por seleção em população de milho pipoca de quinto ciclo de seleção recorrente. *Revista Ciência Agronômica*, 42(2): 473-481.
- RANGEL R.M., Amaral Júnior A.T., Scapim C.A., Freitas Júnior S., Pereira M.G. (2008) Genetic parameters in parents and hybrids of circulant diallel in popcorn. *Genetics and Molecular Research*, 7(4):1020-1030.
- RESENDE Jr M.F.R., Muñoz P., Resende M.D.V. (2012) Accuracy of genomic selection methods in a standard data set of loblolly pine (*Pinus taeda* L.). *Genetics*, 190:1503–10.
- RESENDE Junior M.F.R. (2014) Genomic selection in plant breeding: predicting breeding values, non-additive effects and application to mate-pair allocation. Dissertation, University of Florida.
- RESENDE M.D., Resende Junior M.F.R.R., Aguiar A.M., Abad J.I.M., Missiaggia A.A., Sansaloni C., Petrolí C., Grattapaglia D. (2010) Computação da seleção genômica ampla (GWS). EMBRAPA Florestas.
- RESENDE M.D.V, Lopes P.S., Silva R.L., Pires I.E. (2008) Seleção genômica ampla (GWS) e maximização da eficiência do melhoramento genético. *Pesquisa Florestal Brasileira*, 56: 63-77.
- RESENDE M.D.V. (2008) Genômica quantitativa e seleção no melhoramento de plantas perenes e animais. Colombo: Embrapa Florestas.
- RIBEIRO R.M., Amaral Júnior A.T., Gonçalves L.S.A., Cândido L.S., Silva T.R.C., Pena G.F. (2012). Genetic progress in the UNB-2U population of popcorn under

recurrent selection in Rio de Janeiro. *Genetics and Molecular Research*, 11: 1417-1423.

RIEDELSCHEIMER C., Technow F., Melchinger A.E. (2012) Comparison of whole-genome prediction models for traits with contrasting genetic architecture in a diversity panel of maize inbred lines. *BMC Genomics*, 13: 452

ROBERTSON A. A. (1960) Theory of limits in artificial selection. *Proc. R. Soc. London Ser. B*, 153: 234-249.

SABA H., Vale V.C., Moret M.A., Miranda J.G.V. (2014) Spatio-temporal correlation networks of dengue in the state of Bahia. *BMC Public Health* 14: 1085

SANTOS F.S., Amaral Junior A.T., Freitas Junior S. P., Rangel, R. M., Pereira, M. G. (2007) Predição de ganhos genéticos por índices de seleção na população de milho pipoca UNB-2U sob seleção recorrente. *Bragantia*, 66 (3): 389-396.

SANTOS F.S., Amaral Júnior A.T.D., Júnior F., de Paiva S., Rangel R.M., Scapim C.A., Mora F. (2008) Genetic gain prediction of the third recurrent selection cycle in a popcorn population. *Acta Scientiarum. Agronomy* 30:651-655.

SAWAZAKI E. (2001) A cultura do milho pipoca no Brasil. *O Agrônomo*, Campinas, 53 (2): 11-13.

SCAPIM C.A., Braccini A.L., Pinto R.J.B., Amaral Júnior A.T., Rodovalho M.A., Silva, R.M., Moterle L.M. (2006) Componentes genéticos de medidas e depressão por endogamia em populações de milho-pipoca. *Ciência Rural* 36:36-41

SCAPIM C.A., Pacheco C.A.P., Amaral Junior A.T., Vieira R.A., Pinto R.J.B., Conrado T.V. (2010) Correlations among yield and popping expansion stability parameters in popcorn. *Eupytica*, 174(2): 209-218

SCAPIM C.A., Pacheco C.A.P., Tonet A., Braccini A.L., Pinto R.J.B. (2002) Análise dialélica e heterose de populações de milho pipoca. *Bragantia*, 61 (3): 219-230

SCHWANTES I.A., Amaral Júnior A.T., Gerhardt I.F.S., Vivas M., Lima e Silva F.H., Kamphorst S.H. (2017) Diallel analysis of resistance to Fusarium ear rot in Brazilian popcorn genotypes. *Trop. plant pathol*, 42: 70

- SENASA Serviço Nacional de Sanidade e Qualidade Alimentar da Argentina (2018). Acessado em 10/03/2019 (<https://www.argentina.gob.ar/senasa/exportaciones-importaciones-y-estadisticas>).
- SILVA A.R., Rêgo E.R., Pessoa M.A.S., Rêgo M.M. (2016) Correlation network analysis between phenotypic and genotypic traits of chili pepper. *Pesquisa Agropecuária Brasileira*. 51(4): 372-377
- SILVA V.Q.R., Amaral Júnior A.T., Gonçalves L.S.A., Freitas Júnior S.P., Candido L.S., Vittorazzi C., Moterle L.M., Vieira R.A., Scapim C.A. (2010) Combining ability of tropical and temperate inbred lines of popcorn. *Genetics and Molecular Research*, 9:1742-1750.
- SOLBERG T.R., Sonesson A., Wooliams J., Meuwissen T.H.E. (2008) Genomic selection using different marker types and density. *Journal of Animal Science*. 86: 2447–2454.
- SOLBERG T.R., Sonesson A.K., WoolliamS J.A., Meuwissen T.H.E. (2009) Reducing dimensionality for prediction of genome-wide breeding values. *Genetics Selections Evolution*, 41(29):1-8.
- TERRA T.F. et al, (2011) Genetic variability in maize and teosinte populations estimated by microsatellites markers. *Cienc. Rural*, 41(2):205-211.
- THAVAMANIKUMAR S., Dolferus R., Thumma B.R. (2015) Comparison of genomic selection models to predict flowering time and spike grain number in two hexaploid wheat doubled haploid populations. *G3: Genes, Genomes, Genetics* 5(10):1991-1998
- TIBSHIRANI, R. (1996) Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society Series B*, 58(1):267–288.
- TUSELL L., Pérez-Rodríguez P., Forni S., Gianola D. (2014) Model averaging for genome-enabled prediction with reproducing kernel Hilbert spaces: a case study with pig litter size and wheat yield. *Journal of Animal Breeding and Genetics* 131(2):105-115

- UTSUNOMIYA Y.T., do Carmo A.S., Carvalheiro R., Neves H.H.R., Matos M.C., Zavarez L.B. (2013) Genome-wide association study for birth weight in Nellore cattle points to previously described orthologous genes affecting human and bovine height. *BMC Genetics* 14(52)
- VALENTE M.S.F. et al, (2016) Seleção genômica para melhoramento vegetal com diferentes estruturas populacionais. *Pesq. agropec. bras.* 51(11):1857-1867.
- WANG X., Xua Y., Hu Z., Hu C. (2018) Genomic selection methods for crop improvement: Current status and prospects. *The Crop Journal* 6(4):330-340
- WANG X., Yang Z., Xu C. (2015) A comparison of genomic selection methods for breeding value prediction. *Science Bulletin* 60:925–935
- WEISS K.M., Clark A.G. (2002) Linkage disequilibrium and the mapping of complex human traits. *Trends in Genetics* 18(1):19-24
- WONG C., Bernardo R. (2008) Genomewide selection in oil palm: increasing selection gain per unit time and cost with small populations. *TAG Theoretical and Applied Genetics*, 116:815-824.
- WRAY, N.R., Yang, J., Hayes, B., Price, A., Goddard, M., Visscher, P (2013) “Pitfalls of Predicting Complex Traits from SNPs.” *Nature Reviews. Genetics*, 14(7):507–515.
- XU S., Zhu D., Zhang Q. (2014) Predicting hybrid performance in rice using genomic best linear unbiased prediction. *Proceedings of the National Academy of Sciences of the United States of America*, 111:12456-12461.
- XU Y., Wang X., Ding X., Zheng X., Yang Z., Xu C., Hu Z. (2018) Genomic selection of agronomic traits in hybrid rice using an NCII population. *Rice* 11:32
- YABE S., Ohsawa R., Iwata H. (2013) Potential of genomic selection for mass selection breeding in annual allogamous crops. *Crop Scienc*, 53: 95-105.
- YU J., Buckler E.S. (2006) Genetic association mapping and genome organization of maize. *Curr. Opin. Biotechnol.* 17:155–160.

- ZHANG A., Wang H., Beyene Y., Semagn K., Liu Y., Cao S., Cui Z., Ruan Y., Burgueño J., Vicente F. S., Olsen M., Prasanna B. M., Cruza J., Yu H., Zhang X. (2017) Effect of Trait Heritability, Training Population Size and Marker Density on Genomic Prediction Accuracy Estimation in 22 bi-parental Tropical Maize Populations. *Frontiers in Plant Science*, 8:1916.
- ZHAO Y., Mette M.F., Reif J.C. (2015) Genomic selection in hybrid breeding. *Plant Breeding*, 134:1-10.
- ZHAO Y., Zeng J., Fernando R., Reif J.C. (2013) Genomic prediction of hybrid wheat performance. *Crop Science*, 53:802-810.
- ZINSLY J.R., Machado J.A. (1987) Milho-pipoca. In: Paterniani, E., Viegas, G.P., *Melhoramento e Produção de Milho*. 2. ed. Campinas: Fundação Cargil, p. 413-421.
- ZOU H., Hastie T., (2005) Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society*, 67(2):301–320.