



INSTITUTO FEDERAL DO PIAUÍ – IFPI
LICENCIATURA EM MATEMÁTICA
CAMPUS DE URUÇUI

ADRIANA CARNEIRO LIMA

**ESTUDO DE SIMULAÇÃO EM SELEÇÃO GENÔMICA EM UM CENÁRIO
OLIGOGÊNICO: rr-BLUP x BAYES B**

URUÇUI – PI

2021

INSTITUTO FEDERAL DO PIAUÍ – IFPI
LICENCIATURA EM MATEMÁTICA
CAMPUS DE URUÇUI

ADRIANA CARNEIRO LIMA

**ESTUDO DE SIMULAÇÃO EM SELEÇÃO GENÔMICA EM UM CENÁRIO
OLIGOGÊNICO: rr-BLUP x BAYES B**

Monografia apresentada ao Instituto Federal do Piauí - Campus Uruçuí,
como parte das exigências do curso de Licenciatura em Matemática, para
a obtenção do título de Licenciado em Matemática.

Orientador: Professor Drº Ernandes Guedes Moura

Uruçuí - PI

2021

Dados Internacionais de Catalogação na Publicação (CIP) de acordo com ISBD

Lima, Adriana Carneiro
L732e Estudo de simulação em seleção genômica em um cenário oligogênico :
rr-BLUP x BAYES B / Adriana Carneiro Lima. - 2021.
11 p.: il. color.
Trabalho de Conclusão de Curso (Licenciatura em Matemática) - Instituto
Federal do Piauí, Campus Uruçuí, 2021.
Orientador : Prof. Dr. Ernandes Guedes Moura.
1. Seleção genômica. 2. Predição genômica. 3. Métodos estatísticos -
comparação. I. Título.

CDD - 510

Elaborado por Aldeide Costa dos Santos Sousa CRB 3/1423

DEDICATÓRIA

Dedico este trabalho primeiramente a minha amada filha, Maria Gabriela. Pois, por ela procuro evoluir a cada dia.

Aos meus queridos pais pela dedicação, amor e apoio a mim sempre dados.

A todos aqueles que contribuíram de forma direta ou indireta para o meu crescimento acadêmico e pessoal.

AGRADECIMENTOS

Agradeço primeiramente a Deus, pois por Ele e para Ele são todas as coisas.

À minha mãe, Maria do Socorro por toda dedicação, cuidado, paciência, exemplo, ensinamento, amor e orações a mim oferecidos ao longo da vida. Ao meu pai, Moacyr, pelo exemplo de pessoa honesta e bondosa que é! À minha filha, Maria Gabriela, por entender minha ausência ao longo do meu percurso acadêmico. E por ser o motivo que me mantém viva desde o seu nascimento até os dias de hoje, amo você!

Aos meus irmãos, Nelson, Fabiano, Ernandes, Anny Cristina e Adriano (in memoria) que sempre me apoiaram e ajudaram.

A todos os amigos, que acredito não ser necessário citar nomes, pois cada pessoa do meu pequeno ciclo de amizade sabe do seu lugar em minha vida e de sua importância pra mim. Obrigada por todos os momentos.

Ao meu professor orientador Ernandes Guedes Moura, pelo tempo e ensinamentos em mim investidos e pela contribuição para minha vida acadêmica. A todos os professores do curso em geral, saibam que foram essenciais para minha trajetória acadêmica no curso de Matemática.

As amizades construídas, por todos os momentos compartilhados, a presença de vocês me motivou chegar até aqui e tornou minha trajetória mais alegre. Espero que cada um saiba de sua importância em minha vida. Meu muito obrigada!!!

[...] A minha alucinação

É suportar o dia-a-dia

E meu delírio

É a experiência

Com coisas reais

[...]

Amar e mudar as coisas

Me interessa mais

Amar e mudar as coisas

Amar e mudar as coisas

Me interessa mais

(BELCHIOR,1976)

RESUMO

Com o aumento do número de marcadores SNPs (single nucleotide polymorphism), houve interesse em aplicar métodos estatísticos na seleção genômica, afim de identificar o mérito genético de um indivíduo. Objetivou-se, neste trabalho, comparar o método rr-BLUP com o método Bayes B num cenário simulado oligogênico quanto as suas capacidades preditivas e também poder de detecção de mapear regiões causais. Para isso, foram simuladas uma população F^2 com 300 indivíduos cada, conforme duas herdabilidades (0,1 e 0,7). Um total de 12150 marcadores SNPs, distribuídos em dez grupos de ligações, também foram simulados. O método Bayes B obteve melhor capacidade preditiva no cenário estudado em relação rr-BLUP.

Palavras-chave: Seleção Genômica, Predição Genômica, Arquitetura Genética.

RESUME

With the increase in the number of SNPs (single nucleotide polymorphism) markers, there was interest in applying statistical methods in genomic selection, in order to identify the genetic merit of an individual. The objective of this work was to compare the rr-BLUP method with the Bayes B method in a simulated oligogenic scenario regarding its predictive capabilities and also its detection power to map causal regions. For this, a population F^2 with 300 individuals each was simulated, according to two heritability (0.1 and 0.7). A total of 12,150 SNPs markers, distributed into ten linkage groups, were also simulated. The Bayes B method obtained better predictive capacity in the studied scenario compared to rr-BLUP.

Keywords: Genomic Selection, Genomic Prediction, Genetic Architecture.

SUMÁRIO

1 INTRODUÇÃO.....	5
2 REFERENCIAL TEÓRICO	6
3 MATERIAL E MÉTODOS	7
3.1 Estimativas de efeito de marcador	7
4 RESULTADOS	8
5 DISCUSSÃO	10
REFERÊNCIAS.....	11

1 INTRODUÇÃO

A seleção genômica ampla (GWS) foi inicialmente proposta por (Meuwissen *et al.* 2001)e, essa metodologia tem sido muito utilizada no melhoramento genético. Com a disponibilidade de marcadores de polimorfismo de nucleotídeo único de alta densidade, a seleção genômica tornou-se um importante tópico de pesquisa em melhoramento de plantas e animais. Assim, finalidade da seleção genômica (SG) é prever fenótipos a partir de genótipos (Wang *et al.* 2018).Em outras palavras, consiste em saturar o DNA com marcadores moleculares, prever seus efeitos e, posteriormente, prever adequadamente o valor genético genômico dos indivíduos da população em estudo, objetivando ganho na seleção para as características de interesse.

Para mensurar o mérito genético, geneticistas e outros pesquisadores dessa área, utilizam modelos estatísticos que consigam acomodar um número grande de variáveis regressoras. Há vários modelos estatísticos utilizados para estimar o mérito genético. Entre os métodos mais utilizados, pode-se destacar os métodos Bayes B Meuwissen *et al.*(2001) e rr-BLUP (Endelman 2011). Esses modelos diferem em capacidade preditiva, custo computacional, e capacidade em lidar com diferentes arquiteturas genéticas.

Diante das consolidações expostas, a presente monografia visa responder como ocorre o processo de comparação de dois métodos de seleção genômicas (Bayes B e rr-BLUP), num cenário simulado, em que poucos genes explicam o caráter de interesse (cenário oligogênico)?.

Para responder essa pergunta estabeleceu-se como objetivo geral “comparar dois métodos de seleção genômicas (Bayes B e rr-BLUP), num cenário simulado, em que poucos genes explicam o caráter de interesse (cenário oligogênico)” para assim identificar o método mais curado,ou seja, com menos vícios e de melhor resultado.

2 REFERENCIAL TEÓRICO

A busca de geneticistas e melhoristas para encontrar uma relação funcional entre os caracteres quantitativos de indivíduos (animais e plantas) têm trazido grandes avanços na ciência genômica. Destaca-se o surgimento de marcadores moleculares que permitem quantificar a variabilidade genética existente no DNA (*Deoxyribonucleicacid*).

Com o uso dos marcadores moleculares foi possível o mapeamento de caracteres quantitativos que são controlados por muitos genes de efeitos pequenos, sendo altamente influenciados pelo ambiente. Assim, distribuindo esses marcadores ao longo do genoma, pode-se investigar, além dos efeitos dos principais genes, também suas respectivas localizações, sendo este processo denominado como mapeamento de QTLs (*QuantitativeTrait Loci*) (Lander and Botstein 1989).

Na busca pela interpretação genética da análise genômica ampla e no desenvolvimento de métodos e aperfeiçoamento de modelos, ambos eficazes para contornar problemas relacionados à seleção de covariáveis, tais como a multicolinearidade e/ou multicolinearidade.

O uso de modelos tem possibilitado relacionar matematicamente variáveis, e a partir da utilização de metodologias estatísticas tem sido possível interpretar covariáveis. Dada à infinidade de modelos existentes adotados na GWS, a utilização de modelos tais como, Oligogênico, Poligênico e Infinitesimal tem sido frequente em pesquisas relacionadas à genômica (Hu *et al.* 2012).

3 MATERIAL E MÉTODOS

Utilizou-se o software QGenesJoehanes *et al.*(2008), para simular dez grupos de ligação de tamanho de 120 cM cada e distância média de 0,001 cM no genoma, em 300 indivíduos pertencentes a uma população F_2 , totalizando 12150 marcadores SNP. Seis SNPs foram assumidos como QTL para representar um cenário oligogênico e seus efeitos amostrados de uma distribuição normal com média igual a zero e desvio padrão igual a 3, $N(0;3)$.

3.1 Estimativas de efeito de marcador

O modelo estatístico para estimação de efeito de marcador pode ser escrito como:

$$y_i = \beta + \sum_{j=1}^p Z_{ij} \gamma_j + \varepsilon_i \quad (01)$$

Em que p é o os marcadores, β é o intercepto, γ_j é o efeito do j-ésimo marcador, Z_{ij} é uma matriz que indica o estado genotípico e ε_i é o erro, e segue uma distribuição normal com média zero e variância desconhecida σ^2 .

4 RESULTADOS

Na Figura 1 encontram-se os efeitos dos QTLs simulados (painéis A e B), e os efeitos preditos pelos métodos rr-BLUP (painéis C e D) e Bayes B (painéis E e F). Observa-se que apenas os segmentos contendo grandes QTLs foram mapeados pelos os dois métodos. Os perfis de efeito de SNP do Bayes B apresentou melhor resolução que o rr-BLUP (Figura 1 C:F), inclusive em magnitude de mapear regiões causais.

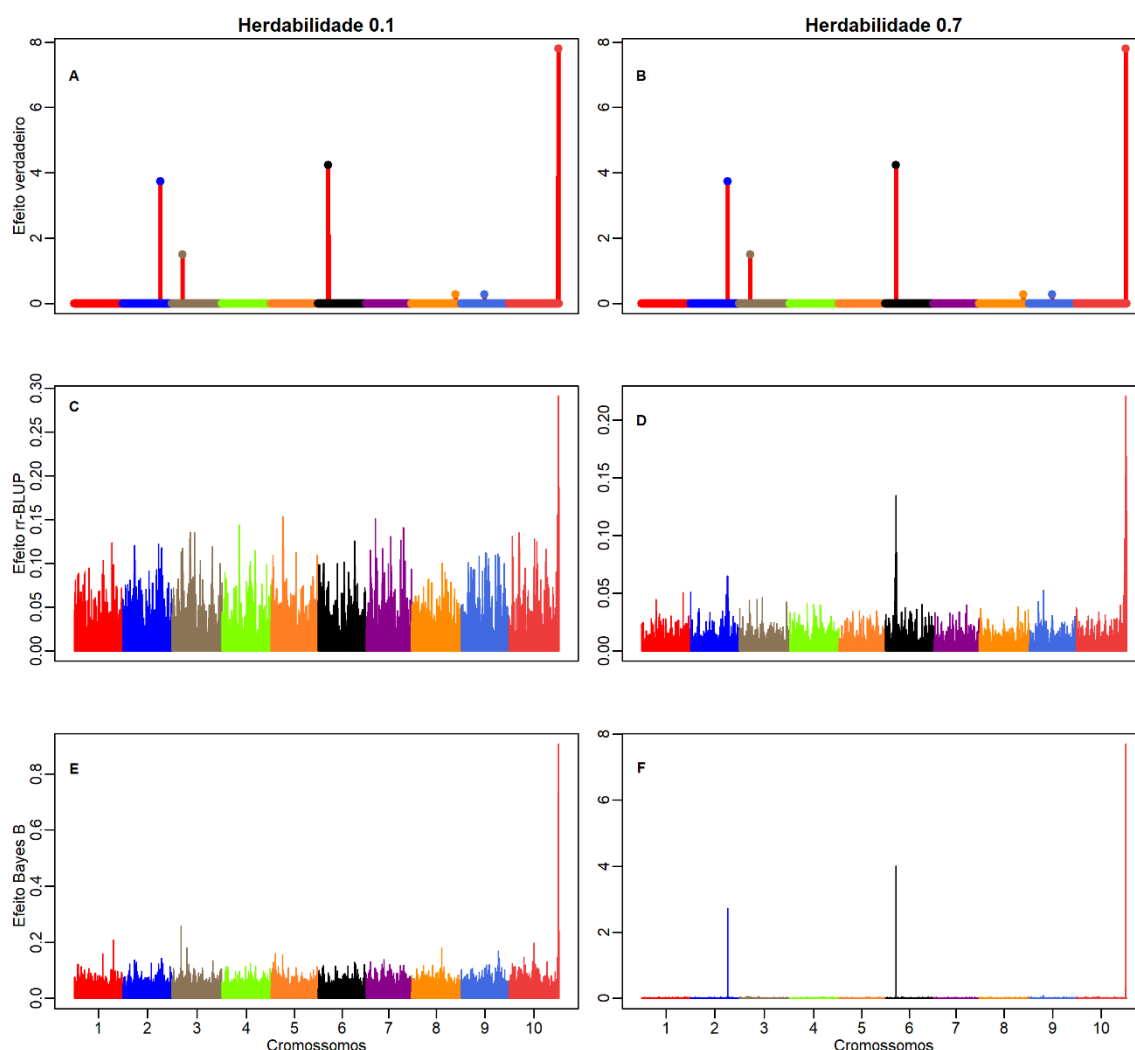


Figura 1. Efeitos simulados verdadeiros de QTL (painel A) ao longo do genoma para as herdabilidades 0,1 e 0,7 e estimados (painéis C:F) a partir dos métodos, respectivamente, rr-BLUP e Bayes B. Pontos coloridos representam os seis verdadeiros QTL distribuídos em 12150 SNP ao longo de dez grupos de ligação.

Nota-se que, com o aumento da herdabilidade, ouve uma melhora expressiva na resolução de ambos os métodos, em especial, o Bays B para herdabilidade alta, obteve melhor resolução.

A Tabela 1 mostra os erros quadráticos médios (EQM) e os coeficientes de determinação (R^2) entre os valores verdadeiros simulados (VGG) e os preditos pelos métodos avaliados, para as duas herdabilidades, nos cenários oligogênico (6 QTL).

Tabela 1. Coeficiente de determinação (R^2) e Erro Quadrático Médio (EQM) entre os valores verdadeiros e preditos, usando os diferentes métodos nos cenários estudados.

População	Modelos	Herdabilidades			
		0,1		0,7	
		EQM	R^2	EQM	R^2
F₂(Oligogênico)	rr-BLUP	6.39	0.23	3.39	0.78
	Bayes B	6.38	0.25	3.24	0.80

Em termos de acurácia, para população F^2 no cenário estudado(oligogênico), o método Bayes B obteve melhores resultados que rr-BLUP, nas duas herdabilidades. Observa-se que com o aumento dos níveis de herdabilidades, os valores de EQM decaem, nos dois métodos.

5 DISCUSSÃO

No presente estudo, realizou-se uma comparação na capacidade preditiva dos métodos Bayes B e rr-BLUP, num cenário oligogênico. Para esse cenário, o método Bayes B foi mais eficiente do que método rr-BLUP. Esses resultados corroboram com os resultados encontrados por (Daetwyler *et al.* 2010). Com os avanços na tecnologia de coleta e armazenamento de dados, foi possível, mensurar o mérito genético e reduzir o tempo para seleção genômica.

As diferenças preditivas no cenário simulado estudado, pode ser o indicativo de que a arquitetura genômica pode influenciar os resultados preditivos. Nesse estudo, o método Bayes B obteve maior capacidade preditiva que rr-BLUP, ao passo que também foi mais eficiente em termo de resolução de mapear regiões causais. No entanto, não se pode afirmar que método Bayes B é melhor que o rr-BLUP, pois pode ter sido beneficiado pelo cenário estudados. De acordo com Moura *et al.* (2019) a escolha do modelo mais adequado para uso na análise do genoma pode depender de vários atributos interconectados, como a arquitetura genética (por exemplo, número de genes). Dessa forma, pode-se afirmar que num cenário oligogênico, o método Bayes B é recomendado tanto em predição quanto para mapear regiões causais.

A método Bayes B funcionou bem para dados simulados de uma população F^2 , num cenário oligogênico. No entanto, não queremos afirmar que esse método seja absolutamente ótimo em outras arquiteturas genéticas (por exemplo, num cenário infinitesimal). Todavia, para o cenário em estudo, além da capacidade preditiva, o grande diferencial do Bayes B foi na alta resolução de mapear regiões causais, obtendo resultados muitos melhores nesse quesito em relação ao concorrente rr-BLUP.

REFERÊNCIAS

Daetwyler H. D., Pong-Wong R., Villanueva B., Woolliams J. A., 2010 The impact of genetic architecture on genome-wide evaluation methods. *Genetics* 185: 1021–1031.

Endelman J. B., 2011 Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. *Plant Genome J.* 4: 250.

Hu Z., Wang Z., Xu S., 2012 An infinitesimal model for quantitative trait genomic value prediction. *PLoS One* 7: 1–14.

Joehanes R., Nelson J. C., Bateman A., 2008 QGene 4.0, an extensible Java QTL-analysis platform. *Bioinforma. Appl. NOTE* 24: 2788–278910.

Lander E. S., Botstein S., 1989 Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 121: 185.

Meuwissen T. H., Hayes B. J., Goddard M. E., 2001 Prediction of total genetic value using genome-wide dense markers maps. *Genetics* 157: 1819–1829.

Moura E. G., Pamplona A. K. A., Balestre M., 2019 Functional models in genome-wide selection (IB Rogozin, Ed.). *PLoS One* 14: 1–27.

Wang X., Xu Y., Hu Z., Xu C., 2018 Genomic selection methods for crop improvement: Current status and prospects. *Crop J.* 6: 330–340.