



MICHELE BARBOSA

**ANÁLISE DE ASSOCIAÇÃO GENÔMICA AMPLA COM
MÚLTIPLAS MARCAS:
SELEÇÃO E ASSOCIAÇÃO GENÔMICA AMPLA UNIFICADAS**

LAVRAS – MG

2019

MICHELE BARBOSA

**ANÁLISE DE ASSOCIAÇÃO GENÔMICA AMPLA COM MÚLTIPLAS MARCAS:
SELEÇÃO E ASSOCIAÇÃO GENÔMICA AMPLA UNIFICADAS**

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, para a obtenção do título de Doutor.

Prof. Dr. Márcio Balestre
Orientador

**LAVRAS – MG
2019**

**Ficha catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca
Universitária da UFLA, com dados informados pelo(a) próprio(a) autor(a).**

Barbosa, Michele

Análise de associação genômica ampla com múltiplas
marcas : seleção e associação genômica ampla unificadas /
Michele Barbosa. – 2019.

206 p. : il.

Orientador: Prof. Dr. Márcio Balestre.

Tese(doutorado)–Universidade Federal de Lavras, 2019.

Bibliografia.

1. Seleção e Associação Genômica. 2. Estrutura Populacional. 3. Estrutura de Parentesco. 4. Modelo Linear Misto. 5. Inferência Bayesiana. I. Balestre, Márcio. II. Título.

MICHELE BARBOSA

**ANÁLISE DE ASSOCIAÇÃO GENÔMICA AMPLA COM MÚLTIPLAS
MARCAS: SELEÇÃO E ASSOCIAÇÃO GENÔMICA AMPLA UNIFICADAS**

**ANALYSIS OF THE GENOME-WIDE ASSOCIATION STUDIES USING
MULTIPLE MARKERS: UNIFIED GENOME-WIDE SELECTION AND
ASSOCIATION STUDIES**

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, para a obtenção do título de Doutor.

APROVADA em 04 de abril de 2019.

| | |
|---------------------------------|----------|
| Prof. Dr. Júlio César Pereira | UFSCAR |
| Prof. Dr. Evandro Novaes | DBI/UFLA |
| Prof. Dr. Tales Jesus Fernandes | DES/UFLA |
| Profa. Dra. Thelma Sáfyadi | DES/UFLA |

Prof. Dr. Márcio Balestre
Orientador

**LAVRAS – MG
2019**

*Por este trabalho ser capaz de transbordar em realizações, que sirva aos outros em honra à
minha mãe. Que seja um trabalho para inventar vida, em honra ao meu pai.
Com carinho especial aos estudiosos de todos os assuntos necessários ao desenvolvimento de
dons. Que o caminho os leve ao estudo constante, iluminando a própria vida e o poder da
alma.
Às mulheres, fonte criativa do universo e grande útero criador, onde tudo o que é semeado faz
florescer,
meus pensamentos especiais e desejos mágicos dedico.*

AGRADECIMENTOS

A Deus pela graça da vida, pela benção em luz, que me fez estar aqui.

À Universidade Federal de Lavras, pela oportunidade de fazer o curso. Ao corpo docente do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária do Departamento de Estatística da Universidade Federal de Lavras (DES / UFLA), com os quais tive o prazer de conviver. Agradeço-lhe pelos conhecimentos, pela experiência, pelo gosto pela pesquisa e pelo apoio. À direção, administração e funcionários, em especial, à Nádya Ferreira, por toda a presteza em fornecer orientações, contribuindo com a qualidade do curso.

À Universidade Federal de Alfenas - UNIFAL-MG, Campus Varginha, pela licença de afastamento para a participação em Programa de pós-graduação. Sou grata a cada membro do corpo docente, à direção e à administração desta instituição de ensino. Reconheço que o trabalho e contribuições profissionais, durante este tempo, fez aprimorar meus estudos, ideias e proposta de elevar o nível profissional, ajudando no desenvolvimento desta instituição.

Ao professor Márcio Balestre, pela sintonia de concordância e respeito, que permitiu reconciliar e fortalecer outros elos, que se conectam ao aprender e servir à vida. Assim, o professor segue comigo naquilo que faço e naquilo que passo aos outros. Como foi feito comigo, darei aos outros o que serve as suas vidas e para as suas realizações.

Aos professores Guilherme Rosa, Júlio Sílvio de Sousa Bueno Filho, Tales Jesus Fernandes e José Airton Rodrigues Nunes pelo apoio na elaboração deste trabalho, pelos pertinentes apontamentos durante a qualificação, que engrandeceram este estudo.

Aos meus pais, Marlene e Antônio, por tomarem um lugar tão especial, serem pais perfeitos exatamente da maneira que o são. Trouxeram-me à vida e nada é maior, portanto, foram muito bem sucedidos, ocupam uma instância abençoada.

Um caminho possível de ajuda só existe porque o vi nos outros seres capazes: Ricardo, Camila e João Pedro. Grata pelos irmãos de minha história e que sempre cederam lugar para muito mais presentes: Igor e Mahê.

Ao Daniel Ferraz Gorgulho, que veio de outro lugar, com culturas, valores e linguagem própria. Grata por ser como é e me dar o que tem dado.

Aos colegas e amigos que me ajudaram, de diferentes formas, em momentos importantes desta trajetória e compartilharam esta experiência comigo: às turmas de mestrado e doutorado, em especial às companheiras de caminhada, Laís Mesquita Silva, Leila Maria Ferreira e Thais Destefani Ribeiro Furtado. Mulheres que se unem em uma mesma busca, conscientes de que o sucesso de uma irmã contribui para fortalecer a todas.

Um agradecimento muito especial aos amigos queridos, Adriana Brasil, Aline Lourenço de Oliveira, Cristian Tiago Erazo Mendes, Danilo Machado Pires, Deive Oliveira, Helena Sanchez Marcon, Josielle Bensi Vicentin Trevisan, Luiz Paulo Miranda Pires, Priscila Damião, Sebastián Fondini por confiarem em mim e estarem ao meu lado em tantos momentos da vida.

Por milhas e milhas, espalhados em várias direções de muitos lugares, avançam meus agradecimentos por esta jornada - a melhor, mensurada em amigos estrangeiros.

Agradeço aos meus alunos, que me encontraram para aprender e acabaram sendo meus mestres.

A mim mesma, por ter persistido nas atitudes que adoto diante da vida. Pela rebeldia, teimosia e loucura de olhar à volta e gostar de tudo que vejo, principalmente por aprender que a realização dos meus sonhos é também daqueles que vieram antes. Assim aconteceu e integro-o em mim, com todo o desafio que isso representa. Ao ter percorrido o caminho e criado novas realidades a partir do sustento da árvore de gerações - sigo, com uma fonte de força surpreendente e plena.

Por fim, esta tese é um compromisso de honra e um processo resultante de várias mãos. Gostaria muito de agradecer às mãos habilidosas, que nos últimos quatro anos, passaram e transpassam possibilidades. Doaram-se na arte de viver bem, para que eu pudesse costurar o "embornal da sabedoria" em que trago as recordações do verdadeiro estudar, erguendo o olhar para contemplar as dádivas presentes em tantos ensinamentos, sintonizados na harmonia.

*"... Quem traria novas flores para a nossa árvore se não fossem as ovelhas negras? Quem criaria novos ramos? Sem elas, os sonhos não realizados daqueles que sustentam a árvore gerações atrás, morreriam enterrados sob as suas próprias raízes. Que ninguém te faça duvidar, cuida da tua "raridade" como a flor mais preciosa da tua árvore. Tu és o sonho de todos os teus antepassados"
(Bert Hellinger).*

RESUMO

As técnicas de Seleção Genômica Ampla (*Genome Wide Selection* - GWS) e Associação Genômica Ampla (*Genome Wide Association Studies* - GWAS) têm sido correntemente aplicadas em interação aos programas de melhoramento genético, sobretudo em culturas agronômicas, o que gera investimentos tecnológico/científico, econômico e humano. Além da eficiência de ambas na seleção e detecção de genes responsáveis por determinadas características fenotípicas, a GWS é uma abordagem experimental que integra ferramentas da genética quantitativa e enfatiza a predição dos efeitos genéticos de milhares de marcadores (SNPs); já o propósito da GWAS é identificar variações na sequência do DNA em todo o genoma associadas a determinadas características fenotípicas de interesse. O presente trabalho pretende unificar, em um mesmo tratamento conceitual, essas duas fontes estratégicas (GWS e GWAS) dos programas de melhoramento genético. Para tanto, implementou-se o estudo de Seleção e Associação Genômica Ampla (*Genome Wide Selection Association Studies* - GWSAS), a partir da abordagem do modelo linear misto, a fim de identificar regiões candidatas a genes a expressar determinada característica, considerando a estrutura populacional. Uma das principais diferenças entre os métodos abordados, para estimar parâmetros e o método requerido para GWSAS diz respeito à caracterização do comportamento das funções de distribuição a posteriori condicional, obtida por meio da eliminação de parâmetros perturbadores. Portanto integrou-se tal função ao método de inferência bayesiana, garantindo a capacidade de modelar as estruturas populacionais e de parentesco, com base na estimativa que capta a dinâmica dos marcadores moleculares (no contexto das microinformações genéticas), ou seja, a estrutura de covariância genômica do núcleo da distribuição a posteriori condicional marginalizada permitiu levantar informações de dependência e de correlação entre os marcadores moleculares, além de apreender intrinsecamente a identidade por estado (*identity by state* - IBS) e a identidade por descendência (*identity by descent* - IBD). Também propôs-se uma solução para inversão da matriz \mathbf{V}_{k-1} , alterando variáveis e produto de vetores, para elevar as facilidades computacionais, no uso das técnicas MCMC; e uma análise aplicada para formalizar a relação entre a GWS (métodos clássicos utilizados: RR-BLUP, Bayes A, Bayes B e BSSV) e a GWAS (com base no modelo proposto por (YU et al., 2006)) com a GWSAS (métodos utilizados: RR-BLUP modificado, Bayes A modificado, Bayes B modificado e BSSV modificado). Assim, exploraram-se os principais aspectos potenciais da GWSAS, enquanto ferramenta atrativa de seleção e associação genômica, ao aplicá-la a 230 linhagens de milho genotipadas, com 23.153 marcadores, a fim de identificar os resultados dos métodos modificados em genes associados a características de interesse, elucidando as propriedades estatísticas e adaptações de cada um, por meio de comparações gráficas e teste de Wald, avaliando os respectivos desempenhos, pelos dados simulados e reais, e considerando a Soma de Quadrados dos Valores Residuais Predito (PRESS), para a herdabilidade e as correlações entre os valores genéticos verdadeiros e preditos. Os resultados foram satisfatórios quanto à capacidade de detecção de genes e baixa taxa de falsos positivos, bem como revelaram melhores índices na seleção mesmo com padrões de contração diferentes. Os perfis dos métodos de seleção foram claramente refletidos em seus respectivos métodos modificados e, conseqüentemente, nos efeitos simulados dos SNPs, sendo considerados confiáveis. Foram propostas ainda análises dos estudos recentes de associação de genômica multiloci, os chamados algoritmos FASTmrMLM e FASTmrEMMA para verificar a capacidade de detectar genes previamente simulados e reais, comparados com todos os outros métodos GWSAS.

Palavras-chave: Seleção e Associação Genômica. Estrutura Populacional. Estrutura de Parentesco. Modelo Linear Misto. Inferência Bayesiana.

ABSTRACT

The Genome-Wide Selection (GWS) and Genome Wide Association Studies (GWAS) techniques have been routinely applied in interaction in breeding programs, especially for agronomic crops, which generates technological/scientific, economic, and human investments. Besides the efficiency of both techniques in selecting and detecting genes responsible for specific phenotypic characteristics, the GWS is an experimental approach that integrates quantitative genetics tools and emphasizes the prediction of the genetic effects of thousands of markers (SNPs). The purpose of the GWAS is to identify variations in the genome-wide DNA sequence associated with specific interesting phenotypic characteristics. The present work intends to unify both breeding programs strategic fronts (GWS and GWAS) in the same conceptual treatment. To do so, we implemented the Genome-Wide Selection Association Studies (GWSAS) using the mixed linear model to identify candidate regions for genes expressing a specific characteristic, considering the structure populational. One of the main differences between the methods used to estimate parameters and the method required for GWSAS is the characterization of the behavior of the conditional a posteriori distribution functions obtained by eliminating disturbing parameters. Therefore, we integrated such a function with the Bayesian inference method, ensuring the capacity to model population and parentage structures based on the estimation that captures the dynamics of molecular markers (in the context of genetic micro-information). In other words, the genomic covariance structure of the nucleus of the marginally conditional a posteriori distribution allowed us to raise the dependence and correlation information between the molecular markers, and intrinsically apprehend the identity by state (IBS) and identity by descent (IBD). We also proposed a solution for inverting the \mathbf{V}_{k-1} matrix, changing variables and vector products to increase the computational facilities when using the MCMC techniques. Furthermore, we applied an analysis to formalize the relationship between the GWS (classical methods used: RR-BLUP, Bayes A, Bayes B, and BSSV) and the GWAS (based on the model proposed by (YU et al., 2006) with the GWSAS (methods used: modified RR-BLUP, modified Bayes A, modified Bayes B, and modified BSSV). Thus, we explored the primary potential aspects of the GWSAS as an attractive tool for genome selection and association by applying it to 230 maize genotyped lines with 23,153 markers to identify the results of the modified methods in genes associated to interesting characteristics, elucidating the statistical properties and adaptations of each one, through graphical comparisons and the Wald test. This methodology allowed us to evaluate the performances by using simulated and real data, while considering the Sum of Squares of the Predicted Residual Values (PRESS) for heritability and the correlations between the real and predicted genetic values. The results were satisfactory concerning the ability to detect genes and low rate of false positives, also revealing better indexes during the selection even with different patterns of contraction. The modified methods and, consequently, the simulated effects of the SNPs, clearly reflected their respective selection method profiles, thus being considered reliable. We propose the analysis of recent multi-loci genomic association studies, the denominated FASTmrMLM and FASTmrEMMA algorithms, to verify the ability to detect previously simulated and real genes, compared to all other GWSAS methods.

Keywords: Genome Selection and Association Studies. Population Structure. Parentage Structure. Linear Mixed Model. Bayesian Inference

LISTA DE FIGURAS

| | |
|---|----|
| Figura 2.1 – Esquema de representação simples para Seleção Genômica | 25 |
| Figura 2.2 – Classificação dos modelos de regressão em Sequenciamento Genômico . . . | 26 |
| Figura 2.3 – Diagrama de identificação de variantes genéticas que distinguem populações de interesse e controle. A análise estatística GWAS calcula a associação entre um SNP para caso interesse/controle, referente ao $-\log_{10}(\text{valor} - p)$ para cada SNP | 34 |
| Figura 2.4 – Representação dos elementos essenciais em GWAS | 36 |
| Figura 3.1 – Arquivo de dados simulados pelo Programa GENES | 67 |
| Figura 3.2 – Painéis demonstrando as marcas dos efeitos simulados em cada cenário Oligogênico (6 genes) e Poligênico (60 genes) | 70 |
| Figura 4.1 – Gráfico de dispersão dos três primeiros componentes principais mostra a distribuição entre as três populações simuladas | 86 |
| Figura 4.2 – <i>Heatmap</i> e dendrograma da matriz de parentesco mostrando nos dados simulados. Vermelho indica alta correlação entre os pares de marcadores e o amarelo baixa correlação | 87 |
| Figura 4.3 – Mapas de calor dos valores de desequilíbrio de ligação (LD) para genoma simulado em cada população. Os marcadores foram ordenados, nos eixos x e y, de modo que cada célula do mapa de calor representa um único par de marcadores. | 88 |
| Figura 4.4 – Cenário Oligogênico - Painéis demonstrando a análise dos efeitos dos marcadores, para cada uma das herdabilidades $h^2 \in \{0,2;0,5;0,8\}$, confrontando o método RR-BLUP para GWS (pacote rrBLUP <i>software</i> R) e método RR-BLUP modificado | 92 |
| Figura 4.5 – Cenário Poligênico - Painéis demonstrando a análise dos efeitos dos marcadores, para cada uma das herdabilidades $h^2 \in \{0,2;0,5;0,8\}$, confrontando o método RR-BLUP para GWS (pacote rrBLUP <i>software</i> R) e método RR-BLUP modificado | 93 |
| Figura 4.6 – Cenário Oligogênico - Painéis demonstrando a análise dos efeitos dos marcadores, para cada uma das herdabilidades $h^2 \in \{0,2;0,5;0,8\}$, confrontando o método Bayes A para GWS (pacote BGLR <i>software</i> R) com o método Bayes A modificado | 94 |

| | |
|--|-----|
| Figura 4.7 – Cenário Poligênico - Painéis demonstrando a análise dos efeitos dos marcadores, para cada uma das herdabilidades $h^2 \in \{0,2;0,5;0,8\}$, confrontando o método Bayes A para GWS (pacote BGLR <i>software</i> R), com o método Bayes A modificado | 95 |
| Figura 4.8 – Cenário Oligogênico - Painéis demonstrando a análise dos efeitos dos marcadores, para cada uma das herdabilidades $h^2 \in \{0,2;0,5;0,8\}$, confrontando o método Bayes B para GWS (pacote BGLR <i>software</i> R), com o método Bayes B modificado | 98 |
| Figura 4.9 – Cenário Poligênico - Painéis demonstrando a análise dos efeitos dos marcadores, para cada uma das herdabilidades $h^2 \in \{0,2;0,5;0,8\}$, confrontando o método Bayes B para GWS (pacote BGLR <i>software</i> R), com o método Bayes B modificado | 99 |
| Figura 4.10 – Cenário Oligogênico - Painéis demonstrando a análise dos efeitos dos marcadores, para cada uma das herdabilidades $h^2 \in \{0,2;0,5;0,8\}$, confrontando o método BSSV, com o método BSSV modificado | 101 |
| Figura 4.11 – Cenário Poligênico- Painéis demonstrando a análise dos efeitos dos marcadores, para cada uma das herdabilidades $h^2 \in \{0,2;0,5;0,8\}$, confrontando o método BSSV, com o método BSSV modificado | 102 |
| Figura 4.12 – Painéis demonstrando a análise de dispersão entre o valor genético genômico (VGG) e valores genéticos estimados obtidos por $W\hat{a}$, para cada uma das herdabilidades $h^2 \in \{0,2;0,5;0,8\}$, confrontando o método RR-BLUP com o método RR-BLUP modificado, nos cenários oligogênico e poligênico. Estimativas do R^2 ajustado e PRESS entre VGG e $W\hat{a}$ | 106 |
| Figura 4.13 – Painéis demonstrando a análise de dispersão entre o valor genético genômico (VGG) e valores genéticos estimados obtidos por $W\hat{a}$, para cada uma das herdabilidades $h^2 \in \{0,2;0,5;0,8\}$, confrontando o método Bayes A com o método Bayes A modificado, nos cenários oligogênico e poligênico. Estimativas do R^2 ajustado e PRESS entre VGG e $W\hat{a}$ | 107 |

| | |
|---|-----|
| Figura 4.14 – Painéis demonstrando a análise de dispersão entre o valor genético genômico (VGG) e valores genéticos estimados obtidos por $W\hat{a}$, para cada uma das herdabilidades $h^2 \in \{0, 2; 0, 5; 0, 8\}$, confrontando o método Bayes B com o método Bayes B modificado, nos cenários oligogênico e poligênico. Estimativas do R^2 ajustado e PRESS entre VGG e $W\hat{a}$ | 108 |
| Figura 4.15 – Painéis demonstrando a análise de dispersão entre o valor genético genômico (VGG) e valores genéticos estimados obtidos por $W\hat{a}$, para cada uma das herdabilidades $h^2 \in \{0, 2; 0, 5; 0, 8\}$, confrontando o método BSSV com o método BSSV modificado, nos cenários oligogênico e poligênico. Estimativas do R^2 ajustado e PRESS entre VGG e $W\hat{a}$ | 109 |
| Figura 4.16 – Cenário Oligogênico - Painéis demonstrando a análise dos efeitos dos marcadores para cada uma das herdabilidades $h^2 \in \{0, 2; 0, 5; 0, 8\}$, confrontando o método GWAS (Manhattan plot de $-\log_{10}(\text{valor} - p)$), com o método RR-BLUP modificado (Manhattan plot de $-\log_{10}(\text{valor} - p)$ e $\text{valor} - p$ do teste Wald), para associação de 6000 SNPs com total de 6 cromossomos | 110 |
| Figura 4.17 – Cenário Poligênico - Painéis demonstrando a análise dos efeitos dos marcadores para cada uma das herdabilidades $h^2 \in \{0, 2; 0, 5; 0, 8\}$, confrontando o método GWAS (Manhattan plot de $-\log_{10}(\text{valor} - p)$), com o método RR-BLUP modificado (Manhattan plot de $-\log_{10}(\text{valor} - p)$ e $\text{valor} - p$ de teste Wald), para associação de 6000 SNPs com total de 6 cromossomos | 111 |
| Figura 4.18 – Cenário Oligogênico - Painéis demonstrando a análise dos efeitos dos marcadores, para cada uma das herdabilidades $h^2 \in \{0, 2; 0, 5; 0, 8\}$, confrontando o método GWAS (Manhattan plot de $-\log_{10}(\text{valor} - p)$), com o método Bayes A modificado (Manhattan plot de $-\log_{10}(\text{valor} - p)$ e $\text{valor} - p$ de teste Wald), para associação de 6000 SNPs com total de 6 cromossomos | 113 |
| Figura 4.19 – Cenário Poligênico - Painéis demonstrando a análise dos efeitos dos marcadores, para cada uma das herdabilidades $h^2 \in \{0, 2; 0, 5; 0, 8\}$, confrontando o método GWAS (Manhattan plot de $-\log_{10}(\text{valor} - p)$), com o método Bayes A modificado (Manhattan plot de $-\log_{10}(\text{valor} - p)$ e $\text{valor} - p$ de teste Wald), para associação de 6000 SNPs com total de 6 cromossomos | 114 |

Figura 4.20 – Cenário Oligogênico - Painéis demonstrando a análise dos efeitos dos marcadores, para cada uma das herdabilidades $h^2 \in \{0, 2; 0, 5; 0, 8\}$, confrontando o método GWAS (Manhattan plot de $-\log_{10}(\text{valor} - p)$), com o método Bayes B modificado (Manhattan plot de $-\log_{10}(\text{valor} - p)$ e $\text{valor} - p$ de teste Wald), para associação de 6000 SNPs com total de 6 cromossomos . 117

Figura 4.21 – Cenário Poligênico - Painéis demonstrando a análise dos efeitos dos marcadores, para cada uma das herdabilidades $h^2 \in \{0, 2; 0, 5; 0, 8\}$, confrontando o método GWAS (Manhattan plot de $-\log_{10}(\text{valor} - p)$), com o método Bayes B modificado (Manhattan plot de $-\log_{10}(\text{valor} - p)$ e $\text{valor} - p$ de teste Wald), para associação de 6000 SNPs com total de 6 cromossomos . 118

Figura 4.22 – Cenário Oligogênico - Painéis demonstrando a análise dos efeitos dos marcadores, para cada uma das herdabilidades $h^2 \in \{0, 2; 0, 5; 0, 8\}$, confrontando o método GWAS (Manhattan plot de $-\log_{10}(\text{valor} - p)$), com o método BSSV modificado (Manhattan plot de $-\log_{10}(\text{valor} - p)$ e $\text{valor} - p$ de teste Wald), para associação de 6000 SNPs com total de 6 cromossomos . 121

Figura 4.23 – Cenário Poligênico - Painéis demonstrando a análise dos efeitos dos marcadores, para cada uma das herdabilidades $h^2 \in \{0, 2; 0, 5; 0, 8\}$, confrontando o método GWAS (Manhattan plot de $-\log_{10}(\text{valor} - p)$), com o método BSSV modificado (Manhattan plot de $-\log_{10}(\text{valor} - p)$ e $\text{valor} - p$ de teste Wald), para associação de 6000 SNPs com total de 6 cromossomos . 122

Figura 4.24 – Cenário Oligogênico - Painéis demonstrando a análise dos efeitos dos marcadores, para cada uma das herdabilidades $h^2 \in \{0, 2; 0, 5; 0, 8\}$, confrontando o método FASTmrEMMA e FASTmrMLM (Manhattan plot de $-\log_{10}(\text{valor} - p)$), com o método RR-BLUP modificado (Manhattan plot de $-\log_{10}(\text{valor} - p)$ e $\text{valor} - p$ do teste Wald), para associação de 6000 SNPs com total de 6 cromossomos 125

Figura 4.25 – Cenário Poligênico - Painéis demonstrando a análise dos efeitos dos marcadores, para cada uma das herdabilidades $h^2 \in \{0, 2; 0, 5; 0, 8\}$, confrontando o método FASTmrEMMA e FASTmrMLM (Manhattan plot de $-\log_{10}(\text{valor} - p)$), com o método RR-BLUP modificado (Manhattan plot de $-\log_{10}(\text{valor} - p)$ e $\text{valor} - p$ de teste Wald), para associação de 6000 SNPs com total de 6 cromossomos 126

| | |
|---|-----|
| Figura 4.26 – Cenário Oligogênico - Painéis demonstrando a análise dos efeitos dos marcadores, para cada uma das herdabilidades $h^2 \in \{0, 2; 0, 5; 0, 8\}$, confrontando o método FASTmrEMMA e FASTmrMLM (Manhattan plot de $-\log_{10}(\text{valor} - p)$), com o método Bayes A modificado (Manhattan plot de $-\log_{10}(\text{valor} - p)$ e $\text{valor} - p$ do teste Wald), para associação de 6000 SNPs com total de 6 cromossomos | 127 |
| Figura 4.27 – Cenário Poligênico - Painéis demonstrando a análise dos efeitos dos marcadores, para cada uma das herdabilidades $h^2 \in \{0, 2; 0, 5; 0, 8\}$, confrontando o método FASTmrEMMA e FASTmrMLM (Manhattan plot de $-\log_{10}(\text{valor} - p)$), com o método Bayes A modificado (Manhattan plot de $-\log_{10}(\text{valor} - p)$ e $\text{valor} - p$ de teste Wald), para associação de 6000 SNPs com total de 6 cromossomos | 128 |
| Figura 4.28 – Cenário Oligogênico - Painéis demonstrando a análise dos efeitos dos marcadores, para cada uma das herdabilidades $h^2 \in \{0, 2; 0, 5; 0, 8\}$, confrontando o método FASTmrEMMA e FASTmrMLM (Manhattan plot de $-\log_{10}(\text{valor} - p)$), com o método Bayes B modificado (Manhattan plot de $-\log_{10}(\text{valor} - p)$ e $\text{valor} - p$ do teste Wald), para associação de 6000 SNPs com total de 6 cromossomos | 129 |
| Figura 4.29 – Cenário Poligênico - Painéis demonstrando a análise dos efeitos dos marcadores, para cada uma das herdabilidades $h^2 \in \{0, 2; 0, 5; 0, 8\}$, confrontando o método FASTmrEMMA e FASTmrMLM (Manhattan plot de $-\log_{10}(\text{valor} - p)$), com o método Bayes B modificado (Manhattan plot de $-\log_{10}(\text{valor} - p)$ e $\text{valor} - p$ de teste Wald), para associação de 6000 SNPs com total de 6 cromossomos | 130 |
| Figura 4.30 – Cenário Oligogênico - Painéis demonstrando a análise dos efeitos dos marcadores, para cada uma das herdabilidades $h^2 \in \{0, 2; 0, 5; 0, 8\}$, confrontando o método FASTmrEMMA e FASTmrMLM (Manhattan plot de $-\log_{10}(\text{valor} - p)$), com o método BSSV modificado (Manhattan plot de $-\log_{10}(\text{valor} - p)$ e $\text{valor} - p$ de teste Wald), para associação de 6000 SNPs com total de 6 cromossomos | 131 |

| | |
|---|-----|
| Figura 4.31 – Cenário Poligênico - Painéis demonstrando a análise dos efeitos dos marcadores, para cada uma das herdabilidades $h^2 \in \{0, 2; 0, 5; 0, 8\}$, confrontando o método FASTmrEMMA e FASTmrMLM (Manhattan plot de $-\log_{10}(\text{valor} - p)$), com o método BSSV modificado (Manhattan plot de $-\log_{10}(\text{valor} - p)$ e $\text{valor} - p$ de teste Wald), para associação de 6000 SNPs com total de 6 cromossomos | 132 |
| Figura 4.32 – <i>Heatmap</i> e dendrograma das 230 linhagens da matriz de parentesco mostrando diferença na expressão gênica. Vermelho indica alta correlação entre os pares de marcadores e o amarelo, baixa correlação | 134 |
| Figura 4.33 – Painéis demonstrando a análise dos efeitos dos marcadores, confrontando os métodos RR-BLUP, Bayes A, Bayes B e BSSV, com os métodos RR-BLUP modificado, Bayes A modificado, Bayes B modificado e BSSV modificado, para associação de 23153 SNPs | 136 |
| Figura 4.34 – Painéis demonstrando a análise dos efeitos dos marcadores, confrontando o método GWAS (Manhattan plot de $-\log_{10}(\text{valor} - p)$), com os métodos RR-BLUP modificado, Bayes A modificado, Bayes B modificado e BSSV modificado (Manhattan plot de $-\log_{10}(\text{valor} - p)$ e $\text{valor} - p$ de teste Wald), para associação de 23153 SNPs | 137 |
| Figura 4.35 – Painéis demonstrando a análise dos efeitos dos marcadores, confrontando o método FASTmrMLM (Manhattan plot de $-\log_{10}(\text{valor} - p)$), com os métodos RR-BLUP modificado, Bayes A modificado, Bayes B modificado e BSSV modificado (Manhattan plot de $-\log_{10}(\text{valor} - p)$ e $\text{valor} - p$ de teste Wald), para associação de 23153 SNPs | 138 |
| Figura 1 – Estimativas da herdabilidade (h^2), variância genética (σ_a^2) e variância residual (σ_e^2) para vetores observação simulados em GWAS nos cenários oligogênico e poligênico pelo pacote GAPIT | 161 |
| Figura 2 – Gráficos QQ-plot pelo pacote rrBLUP para GWAS nos cenários oligogênico e poligênico | 162 |
| Figura 3 – Gráfico Manhattan (a) e QQ-plot (b) pela função GWAS pacote rrBLUP | 203 |

LISTA DE TABELAS

| | |
|--|-----|
| Tabela 2.1 – Principais características de alguns métodos preditivos utilizados em Seleção Genômica | 27 |
| Tabela 4.1 – Estimativas da variância genética aditiva (σ_a^2) e residual (σ_e^2) para cada vetor fenotípico simulado, considerando herdabilidades $h^2 \in \{0,2;0,5;0,8\}$ | 88 |
| Tabela 4.2 – Estatísticas das três populações distintas em cada cenário | 89 |
| Tabela 4.3 – Estatísticas das três populações distintas após incorporação do resíduo para cada fenótipo | 89 |
| Tabela 4.4 – Resumo dos resultados de seleção de modelo pelo critério de informação bayesiano (BIC) ajustado pelo pacote GAPIT | 90 |
| Tabela 4.5 – Número de marcadores com valores de efeito relativamente moderados (ou maiores), detectados pelos diferentes métodos, dada a associação ao gene de interesse nos cenários oligogênico (6 genes simulados) e cenário poligênico (60 genes simulados) | 103 |
| Tabela 4.6 – Estimativas do coeficiente de herdabilidade \hat{h}^2 para cada um dos métodos nos cenários oligogênico e poligênico | 104 |
| Tabela 4.7 – Detecção das marcas significativas por cromossomo. Resultados de associação genômica GWAS pacote rrBLUP | 111 |
| Tabela 4.8 – Detecção das marcas significativas por cromossomo. Resultados de associação genômica GWAS pacote rrBLUP e Método Bayes A modificado . . . | 115 |
| Tabela 4.9 – Detecção das marcas significativas por cromossomo. Resultados de associação genômica GWAS pacote rrBLUP e Método Bayes B modificado . . . | 119 |
| Tabela 4.10 – Detecção das marcas significativas por cromossomo. Resultados de associação genômica GWAS pacote rrBLUP e Método BSSV modificado | 123 |
| Tabela 4.11 – Estimativas da variância genética aditiva (σ_a^2) e residual (σ_e^2) e estimativa da herdabilidade (h^2) para análise conjunta de Lavras e Uberlândia | 133 |
| Tabela 4.12 – Resumo dos resultados de seleção de modelo pelo critério de informação bayesiano (BIC) | 134 |
| Tabela 4.13 – Estimativas do coeficiente de herdabilidade \hat{h}^2 , R^2 e PRESS para cada um dos métodos | 135 |

| | |
|--|-----|
| Tabela 4.14 – Detecção das marcas significativas por cromossomo. Resultados de associação genômica FASTmrMLM pacote mrMLM, para associação de 23153 SNPs | 138 |
| Tabela 1 – Tabelas Efeitos Simulados | 163 |
| Tabela 2 – RR-BLUP × RR-BLUP modificado cenário Oligogênico, $h^2 = 0,2$ | 164 |
| Tabela 3 – RR-BLUP × RR-BLUP modificado cenário Oligogênico, $h^2 = 0,5$ | 164 |
| Tabela 4 – RR-BLUP × RR-BLUP modificado cenário Oligogênico, $h^2 = 0,8$ | 164 |
| Tabela 5 – RR-BLUP × RR-BLUP modificado cenário Poligênico, $h^2 = 0,2$ | 165 |
| Tabela 6 – RR-BLUP × RR-BLUP modificado cenário Poligênico, $h^2 = 0,5$ | 166 |
| Tabela 7 – RR-BLUP × RR-BLUP modificado cenário Poligênico, $h^2 = 0,8$ | 167 |
| Tabela 8 – Bayes A × Bayes A modificado cenário Oligogênico, $h^2 = 0,2$ | 168 |
| Tabela 9 – Bayes A × Bayes A modificado cenário Oligogênico, $h^2 = 0,5$ | 168 |
| Tabela 10 – Bayes A × Bayes A modificado cenário Oligogênico, $h^2 = 0,8$ | 168 |
| Tabela 11 – Bayes A × Bayes A modificado cenário Poligênico, $h^2 = 0,2$ | 169 |
| Tabela 12 – Bayes A × Bayes A modificado cenário Poligênico, $h^2 = 0,5$ | 170 |
| Tabela 13 – Bayes A × Bayes A modificado cenário Poligênico, $h^2 = 0,8$ | 171 |
| Tabela 14 – Bayes B × Bayes B modificado cenário Oligogênico, $h^2 = 0,2$ | 172 |
| Tabela 15 – Bayes B × Bayes B modificado cenário Oligogênico, $h^2 = 0,5$ | 172 |
| Tabela 16 – Bayes B × Bayes B modificado cenário Oligogênico, $h^2 = 0,8$ | 172 |
| Tabela 17 – Bayes B × Bayes B modificado cenário Poligênico, $h^2 = 0,2$ | 173 |
| Tabela 18 – Bayes B × Bayes B modificado cenário Poligênico, $h^2 = 0,5$ | 174 |
| Tabela 19 – Bayes B × Bayes B modificado cenário Poligênico, $h^2 = 0,8$ | 175 |
| Tabela 20 – BSSV × BSSV modificado cenário Oligogênico, $h^2 = 0,2$ | 176 |
| Tabela 21 – BSSV × BSSV modificado cenário Oligogênico, $h^2 = 0,5$ | 176 |
| Tabela 22 – BSSV × BSSV modificado cenário Oligogênico, $h^2 = 0,8$ | 176 |
| Tabela 23 – BSSV × BSSV modificado cenário Poligênico, $h^2 = 0,2$ | 177 |
| Tabela 24 – BSSV × BSSV modificado cenário Poligênico, $h^2 = 0,5$ | 178 |
| Tabela 25 – BSSV × BSSV modificado cenário Poligênico, $h^2 = 0,8$ | 179 |
| Tabela 26 – RR-BLUP modificado × GWAS cenário Oligogênico, $h^2 = 0,2$ | 180 |
| Tabela 27 – RR-BLUP modificado × GWAS cenário Oligogênico, $h^2 = 0,5$ | 180 |
| Tabela 28 – RR-BLUP modificado × GWAS cenário Oligogênico, $h^2 = 0,8$ | 180 |
| Tabela 29 – RR-BLUP modificado × GWAS cenário Poligênico, $h^2 = 0,2$ | 181 |

| | |
|---|-----|
| Tabela 30 – RR-BLUP modificado × GWAS cenário Poligênico, $h^2 = 0,5$ | 182 |
| Tabela 31 – RR-BLUP modificado × GWAS cenário Poligênico, $h^2 = 0,8$ | 183 |
| Tabela 32 – Bayes A modificado × GWAS cenário Oligogênico, $h^2 = 0,2$ | 184 |
| Tabela 33 – Bayes A modificado × GWAS cenário Oligogênico, $h^2 = 0,5$ | 184 |
| Tabela 34 – Bayes A modificado × GWAS cenário Oligogênico, $h^2 = 0,8$ | 184 |
| Tabela 35 – Bayes A modificado × GWAS cenário Poligênico, $h^2 = 0,2$ | 185 |
| Tabela 36 – Bayes A modificado × GWAS cenário Poligênico, $h^2 = 0,5$ | 186 |
| Tabela 37 – Bayes A modificado × GWAS cenário Poligênico, $h^2 = 0,8$ | 187 |
| Tabela 38 – Bayes B modificado × GWAS cenário Oligogênico, $h^2 = 0,2$ | 188 |
| Tabela 39 – Bayes B modificado × GWAS cenário Oligogênico, $h^2 = 0,5$ | 188 |
| Tabela 40 – Bayes B modificado × GWAS cenário Oligogênico, $h^2 = 0,8$ | 188 |
| Tabela 41 – Bayes B modificado × GWAS cenário Poligênico, $h^2 = 0,2$ | 189 |
| Tabela 42 – Bayes B modificado × GWAS cenário Poligênico, $h^2 = 0,5$ | 190 |
| Tabela 43 – Bayes B modificado × GWAS cenário Poligênico, $h^2 = 0,8$ | 191 |
| Tabela 44 – BSSV modificado × GWAS cenário Oligogênico, $h^2 = 0,2$ | 192 |
| Tabela 45 – BSSV modificado × GWAS cenário Oligogênico, $h^2 = 0,5$ | 192 |
| Tabela 46 – BSSV modificado × GWAS cenário Oligogênico, $h^2 = 0,8$ | 192 |
| Tabela 47 – BSSV modificado × GWAS cenário Poligênico, $h^2 = 0,2$ | 193 |
| Tabela 48 – BSSV modificado × GWAS cenário Poligênico, $h^2 = 0,5$ | 194 |
| Tabela 49 – BSSV modificado × GWAS cenário Poligênico, $h^2 = 0,8$ | 195 |
| Tabela 50 – FASTmrEMMA cenário Oligogênico | 196 |
| Tabela 51 – FASTmrEMMA cenário Poligênico | 197 |
| Tabela 52 – FASTmrMLM cenário Oligogênico | 198 |
| Tabela 53 – FASTmrMLM cenário Poligênico | 199 |
| Tabela 54 – Distribuições a priori disponíveis para coeficientes de regressão no pacote BGLR | 201 |
| Tabela 55 – Distribuições a priori implementadas no pacote BGLR | 201 |
| Tabela 56 – Comparação de alguns métodos para GWAS | 206 |

SUMÁRIO

| | | |
|-----------|--|----|
| 1 | INTRODUÇÃO | 20 |
| 2 | REFERENCIAL TEÓRICO | 24 |
| 2.1 | Seleção Genômica Ampla (GWS) | 24 |
| 2.1.1 | Modelos Usuais de Seleção Genômica | 28 |
| 2.1.1.1 | RR-BLUP | 28 |
| 2.1.1.2 | Bayes A | 29 |
| 2.1.1.3 | Bayes B | 30 |
| 2.1.2 | BSSV | 31 |
| 2.2 | Associação Genômica Ampla (GWAS) | 33 |
| 2.2.1 | Métodos Paramétricos para GWAS | 35 |
| 2.2.1.1 | Análise de Associação Único Locus | 36 |
| 2.2.1.1.1 | Matriz de Parentesco Genômico | 37 |
| 2.2.1.1.2 | Testes Bonferroni e FDR | 39 |
| 2.2.1.2 | Análise de Associação Multiloci | 41 |
| 2.2.1.3 | Desenvolvimento teórico gerais sobre as diferenças entre GWS e GWAS | 44 |
| 2.3 | O Modelo Linear Misto | 47 |
| 2.3.1 | Estimação por ML e REML | 49 |
| 2.3.2 | Solução Efeito Fixo e Aleatório | 51 |
| 2.4 | Inferência Bayesiana | 53 |
| 2.4.1 | Uso da inferência Bayesiana para a marginalização de parâmetros perturbadores | 55 |
| 2.4.2 | Modelagem Hierárquica | 56 |
| 2.4.3 | Técnica Computacional | 57 |
| 2.5 | Formulação Modelo Múltiplas Marcas para GWAS | 58 |
| 2.5.1 | Recuperação da Informação de Parentesco e Inversão da matriz V_2 | 63 |
| 3 | MATERIAL E MÉTODOS | 66 |
| 3.1 | Material | 66 |
| 3.1.1 | Simulação do Genótipo | 66 |
| 3.1.2 | Simulação dos Fenótipos | 68 |
| 3.1.3 | Dados de resistência à podridão de espiga causada por <i>Fusarium verticillioides</i> | 71 |

| | | |
|----------|---|-----|
| 3.2 | Métodos | 71 |
| 3.2.1 | Modelo Linear Misto para Análise Bayesiana | 71 |
| 3.2.2 | Distribuições a Priori | 72 |
| 3.2.3 | Distribuição a Posteriori Conjunta | 73 |
| 3.2.4 | Marginalização dos Parâmetros Perturbadores | 74 |
| 3.2.4.1 | Inversão da Matriz V_{k-1} | 76 |
| 3.2.5 | Métodos de Análise para Seleção e Associação Genômica Ampla | 77 |
| 3.2.5.1 | RR-BLUP Modificado | 77 |
| 3.2.5.2 | Bayes A Modificado | 79 |
| 3.2.5.3 | Bayes B Modificado | 80 |
| 3.2.5.4 | BSSV Modificado | 81 |
| 3.2.5.5 | Implementação dos Métodos em R | 82 |
| 3.2.6 | Teste de Wald | 83 |
| 3.3 | Critérios de Comparação | 84 |
| 4 | RESULTADOS E DISCUSSÃO | 86 |
| 4.1 | Diagnóstico Genotípico | 86 |
| 4.1.1 | Análise do Desequilíbrio de Ligação | 86 |
| 4.1.2 | Diagnóstico Fenotípico | 88 |
| 4.1.3 | Análise Comparativa dos Métodos | 91 |
| 4.1.3.1 | Análise Comparativa RR-BLUP versus RR-BLUP modificado | 92 |
| 4.1.3.2 | Análise Comparativa Bayes A versus Bayes A modificado | 94 |
| 4.1.3.3 | Análise Comparativa Bayes B versus Bayes B modificado | 98 |
| 4.1.3.4 | Análise Comparativa BSSV versus BSSV modificado | 101 |
| 4.1.3.5 | Análise Comparativa RR-BLUP modificado versus GWAS | 110 |
| 4.1.3.6 | Análise Comparativa Bayes A modificado versus GWAS | 113 |
| 4.1.3.7 | Análise Comparativa Bayes B modificado versus GWAS | 117 |
| 4.1.3.8 | Análise Comparativa BSSV modificado versus GWAS | 121 |
| 4.1.3.9 | Análise Comparativa RR-BLUP modificado versus FASTmrEMMA FASTmrMLM | 125 |
| 4.1.3.10 | Análise Comparativa Bayes A modificado versus FASTmrEMMA e FASTmrMLM | 127 |
| 4.1.3.11 | Análise Comparativa Bayes B modificado versus FASTmrEMMA FASTmrMLM | 129 |

| | | |
|-----------------|--|------------|
| 4.1.3.12 | Análise Comparativa BSSV modificado versus FASTmrMLM e FASTmrEMMA | 131 |
| 4.2 | Comparação dos métodos propostos em aplicação a dados de resistência à podridão da espiga causada por <i>Fusarium verticillioides</i> | 133 |
| 4.3 | Considerações Finais | 139 |
| 5 | CONCLUSÃO | 141 |
| | REFERÊNCIAS | 142 |
| | APÊNDICE A – MODELO MÚLTIPLAS MARCAS PARA GWAS | 148 |

1 INTRODUÇÃO

Com o avanço da bioinformática pela abundância de marcadores gerados, como os SNPs (do inglês *Single Nucleotide Polimorphism*), técnicas para seleção e associação genômica passaram a ser amplamente usadas em culturas agrônomicas. Essas técnicas representaram poderosas ferramentas para análise de características genéticas relacionadas à diferenciação das espécies vegetais.

A maior disponibilidade e credibilidade de informações, ao se realizar varreduras de genomas completos (número de marcadores SNPs superior ao número de indivíduos), possibilitou proceder com mais incentivos tanto em Análises de Seleção Genômica Ampla (*Genome Wide Selection* - GWS) quanto em Análises de Associação Genômica Ampla (*Genome Wide Association* - GWAS). Separadas conceitualmente, essas análises trouxeram desafios referentes à demanda e utilização dos métodos, o que implica a proposta de novas perspectivas que permitam a integração de suas metodologias, para melhor entendimento da arquitetura genética e apontem estratégias de desenvolvimento para análises de seleção e associação genômica.

Idealizada por Meuwissen, Hayes e Goddard (2001), a abordagem inicial de seleção genômica utiliza um grande número de marcadores para predizer fenótipos em uma população. Os métodos bayesianos (Bayes A e Bayes B) – assumem o efeito dos marcadores independentes e produzem uma grande quantidade de marcadores SNPs associados a efeitos nulos – ganharam considerável popularidade pela maior precisão na predição dos valores reprodutivos e genotípicos, passando a ser usados com sucesso. O método Bayes B abarca o Bayes A, sendo os efeitos dos marcadores característicos de distribuição Gaussiana ou t-Student (Bayes A) ou uma mistura de qualquer uma dessas duas densidades com centro de massa zero (Bayes B).

Outras metodologias relacionadas à seleção genômica foram propostas e aprimoradas, sendo diferenciadas pela pressuposição relativa às variâncias, como *Ridge Regression - Best Linear Unbiased Prediction* RR-BLUP e os métodos bayesianos. A velocidade com que as propostas de análises de seleção foram criadas e a dificuldade em relacionar novos desenvolvimentos à teoria existente passou a ser uma das críticas de Gianola et al. (2009), que apresentaram uma revisão dos conceitos teóricos, em análises bayesianas hierárquicas, envolvendo o chamado alfabeto bayesiano. Gianola (2013), também, esclareceu posteriormente o fato do crescente número de letras do alfabeto ser usado, para denotar várias regressões lineares bayesianas, que diferem nas priori adotadas, enquanto compartilham o mesmo modelo de amostragem.

O uso de análises bayesianas hierárquicas para GWS proliferou particularmente na comunidade genética.

A Associação Genômica Ampla (GWAS) também dispõe da matriz de marcadores moleculares SNPs em alta densidade e é empregada, em vários estudos de genética vegetal e animal, o que possibilita a identificação de loci que influenciam uma característica de interesse, como, por exemplo, aqueles que acometem uma doença. O método mais comum de analisar GWAS é uma regressão simples, para cada SNP genotipado, que tem o resultado corrigido por testes múltiplos cujos efeitos significativos são identificados.

Contudo os testes estatísticos múltiplos usados para identificar marcas associadas ao fenótipo nas análises GWAS, acarretam problemas de independência por não avaliarem de forma simultânea os efeitos dos marcadores em todo o genoma. Uma solução possível é a estimativa conjunta de marcadores, como na seleção genômica. Porém, como o número de observações é muito menor que o número de marcadores, utilizam-se restrições como função penalizadora, o que pode apresentar consideráveis desvantagens nos ajustes iterativos, considerando um grande número de parâmetros.

Segundo Gianola et al. (2016), o aprimoramento dos métodos em GWAS ocorreu, a partir do uso de modelos lineares mistos, originalmente desenvolvidos em criação de animais por Henderson (1984), com o propósito de verificar a correlação das observações por semelhanças genotípicas. Ao longo destas propostas, a análise apresentada por Yu et al. (2006) como outra possibilidade, para abordagem em GWAS, motivou a identificação de locus específico e integrou o modelo de estrutura populacional e de parentesco como alternativas, para aumentar a influência genética sob a expressão fenotípica, permitindo estudar o comportamento de redução do número de falsos positivos e aumentar o poder de detecção dos QTLs. Um dos desafios apontados pelas análises GWAS foi a adequação de modelos que corrigem a matriz de parentesco genômico e a estrutura populacional, para realizarem inferências sobre regiões genômicas, que apresentam desequilíbrio de ligação (*disequilibrium linkage* - LD) com os marcadores.

Embora outros métodos tenham seguido com avanços eficazes, para os testes em relação à associação genômica, como admitem Gianola, Hospital e Verrier (2013), Kang et al. (2008) e Zhou e Stephens (2012) recomendam cautela ao empregá-los, ressaltando que são incompletos a estudos de associação genômica, uma vez que, na presença de LD, as métricas-padrão usadas podem ser enganosas para avaliar a contribuição de um locus em relação à variância genética.

Os conteúdos de aplicação à GWAS também merecem atenção, porque são os problemas de justificativas de modelo que nem sempre foram identificáveis, como apontam Gianola et al. (2016), salientando que o marcador testado como fixo, em um modelo de regressão que inclui a matriz G (matriz à base de SNPs), também, contribui para a estrutura de covariância, ou seja, o marcador é incluído implicitamente como um efeito aleatório, para o modelo e, assim, acusa contradição. Além disso, no modelo misto, produz-se uma abordagem mal colocada por utilizar a matriz G , sem qualquer alteração quando representando a estrutura da população extraída de componentes principais. Uma alternativa é utilizar métodos Naive GWAS, em que a estrutura da população não é inserida, dado que a matriz G produz toda a informação necessária (SLAVOV et al., 2014).

Os cenários descritos mostram que a GWS e a GWAS são abordadas de maneiras distintas, pois apresentam objetivos divergentes. Em GWS, o alvo é a seleção do valor genético do indivíduo, sem a preocupação de identificar regiões causais e a realização de testes estatísticos; já a GWAS busca identificar genes causais, explorando o desequilíbrio de ligação, a partir de informações da estrutura da população e de *pedigree* entre indivíduos. Embora os enfoques teóricos e metodológicos, em ambas as análises, sejam distintos, é possível desenvolver uma justificativa teórica comum, a fim de obter uma forma de incorporação da estrutura de população, em modelos de múltiplas marcas, bem como se pode explorar a ideia de testes regularizadores para os efeitos aleatórios, como sugerido por Gelman et al. (2014) e Tempelman (2015).

Assim, propusemos uma abordagem que unificou GWS e GWAS, desenvolvendo uma justificativa teórica que pressupõe modelos lineares mistos em situação de modelagem hierárquica bayesiana, para obter uma forma de incorporação da estrutura de população, em modelos de múltiplas marcas, em cenários de populações distintas. A inferência bayesiana assegurou usar amostras dos efeitos dos marcadores, obtidos pela distribuição a posteriori condicional.

O intuito de unificar as abordagens permitiu conduzir particularidades de problemas de colinearidade pela estrutura de população, atingido pela matriz de parentesco genômico, cujo núcleo da informação referiu-se à matriz \mathbf{V}_{k-1} . Com a matriz \mathbf{V}_{k-1} , recuperou-se exatamente o valor de parentesco genômico unificado de Powell e Goddard (2010), no qual, ao considerar todos os loci, a matriz é ponderada pela variância desigual e, ao se assumir variância igual, como em modelos lineares mistos, levou-se em consideração a estrutura da população. Esse resultado ainda assegurou que o marcador ajustado não compartilha informações do efeito poligênico,

o que contempla ações não redundantes de modelos, diferindo das abordagens apresentadas por (YU et al., 2006). Com isso, a representatividade da dinâmica dos marcadores codifica e incorpora informações de independência e de correlação entre os marcadores no seu estado mais interno (microinformações genéticas).

Uma ferramenta para a solução da inversão da matriz \mathbf{V}_{k-1} competiu em alterar variáveis e produto de vetores para elevar as facilidades computacionais no uso das técnicas MCMC, a fim de eliminar a necessidade de transformação em componentes principais.

Também consideramos uma investigação, para formalizar a relação entre as técnicas GWS (métodos clássicos utilizados: RR-BLUP, Bayes A, Bayes B e BSSV) e GWAS com base no modelo proposto por Yu et al. (2006), com os métodos estabelecidos pela técnica GWSAS (métodos utilizados: RR-BLUP modificado, Bayes A modificado, Bayes B modificado e BSSV modificado), para identificar os genes associados a características de interesse - em um total de 230 linhagens de milho genotipadas, com 23.153 marcadores - além de reconhecer e destacar as diferenças em cada método e examinar as correspondências em adequação a cada modelo. As propriedades estatísticas e adaptações de cada um foram elucidadas, por meio de comparações gráficas, teste de Wald, avaliando os respectivos desempenhos, pelos dados simulados e reais.

Foram propostas ainda análises dos estudos recentes de associação de genômica multi-loci, os chamados algoritmos FASTmrMLM e FASTmrEMMA Wang et al. (2016) e Wen et al. (2017), para verificar a capacidade de detectar genes previamente simulados e reais, comparados com todos os outros métodos GWSAS.

2 REFERENCIAL TEÓRICO

2.1 Seleção Genômica Ampla (GWS)

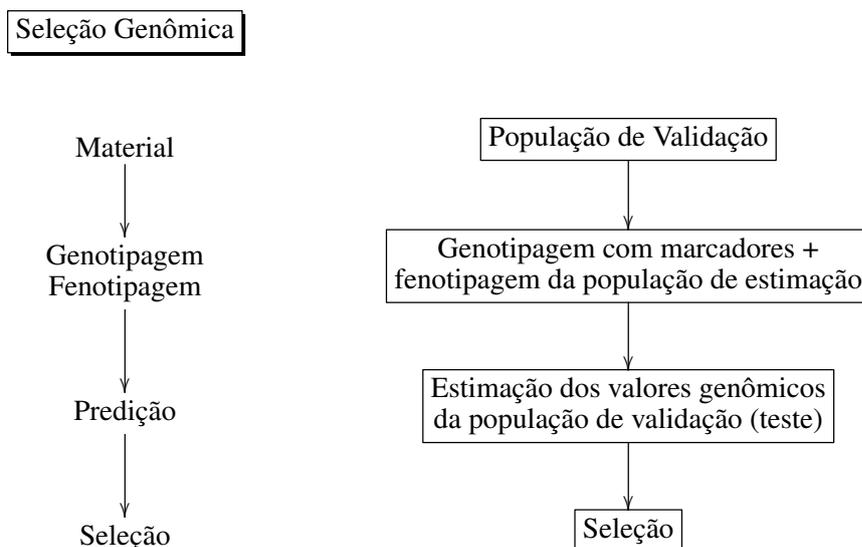
O processo de GWS combina métodos estatísticos com dados genômicos para prever o valor genético em relação à determinada característica. Um pressuposto básico para os estudos de genômica, é que a característica avaliada pode ser explicada, a partir de marcadores genéticos (dentre eles, os SNPs, *Single Nucleotide Polimorphism*). Como nos últimos anos a disseminação e avanços nas tecnologias das plataformas de genotipagem de SNPs progrediram, os métodos de seleção genômica passaram a ser de grande interesse, porém trouxeram incertezas consideráveis na determinação do método mais apropriado para avaliações genômicas (DAETWYLER et al., 2010).

O método de seleção genômica utiliza duas populações: população estimação (ou treinamento - *training population*) e população validação (ou reprodução - *breeding population*), com o objetivo de avaliação genética. De posse das informações genotípicas, a partir dos marcadores, os efeitos são estimados, baseando-se em dados fenotípicos de uma população de estimação. Uma vez estimados os efeitos marcadores, eles são testados na população de validação e, então, são selecionados os marcadores que explicam grande parte da variância genética do caráter em estudo, para que sua informação seja efetivamente incorporada à etapa de seleção, com base no valor genético genômico, alcançando a melhoria desejada (SINGH; SINGH, 2015) (ver Figura 2.1).

O desequilíbrio de ligação (LD) é uma medida do grau de associação entre alelos em uma população. Entre os princípios da GWS, o que define a acurácia da seleção é a existência forte e precisa do desequilíbrio de ligação (LD), em alta resolução, em todo o sequenciamento genômico. O LD é frequentemente usado, para determinar o número ideal de marcadores necessários, para seleção genômica, além de compreender os acontecimentos demográficos passados e evolutivos das populações (HAYES; GODDARD et al., 2001).

Pelo método de seleção genômica ampla descrito por Meuwissen, Hayes e Goddard (2001), com a proposta de aumentar a eficiência dos programas de seleção genômica, tornou-se possível a utilização de um grande número de marcadores amplamente distribuídos, ao longo do genoma, de forma especializada. A acurácia nas estimativas, com relação ao valor genético genômico, para os dados de reprodução animal, ocasionou mudanças importantes na transfor-

Figura 2.1 – Esquema de representação simples para Seleção Genômica



Fonte: Da autora (2019)

mação dos estudos de informação genômica e, com isso, uma variedade de métodos foram sugeridos para estimativas dos valores genéticos.

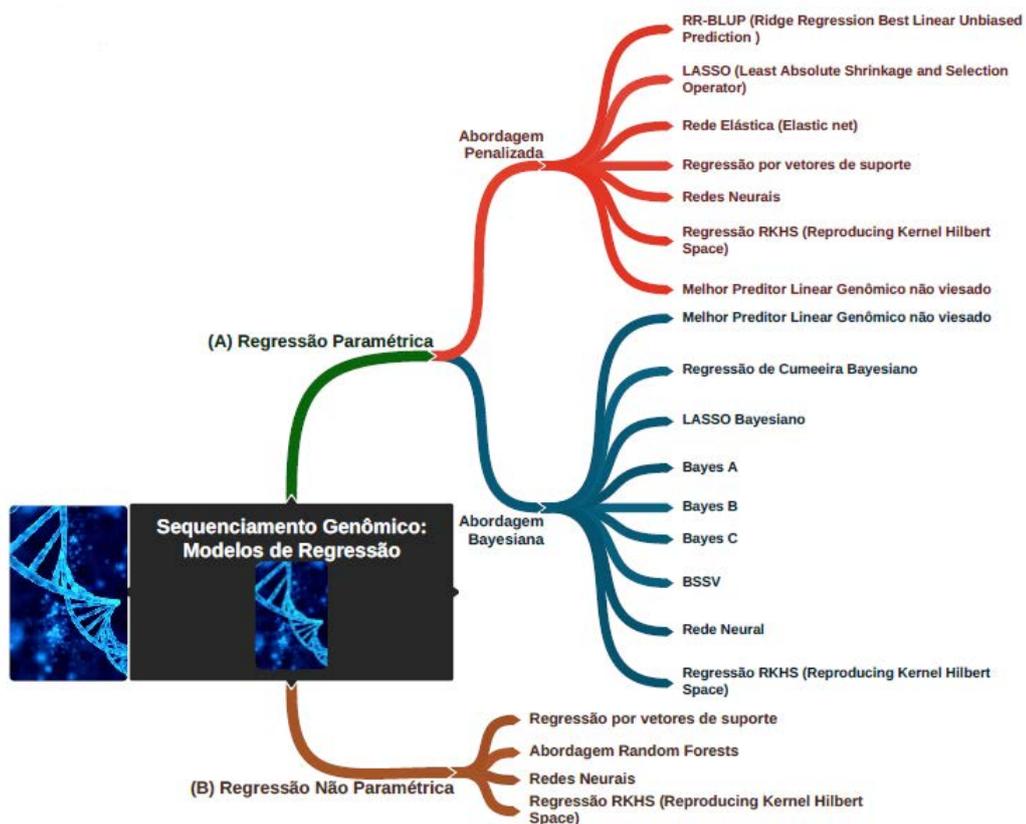
Dentre os métodos iniciais, um primeiro ocupa-se em obter as estimativas do melhor preditor linear não viesado (*best linear unbiased prediction*-BLUP), integra no modelo marcadores cujos efeitos são aleatórios, denominado "*Ridge Regression*" ou "Regressão de Cumeeira", consideram o efeito dos marcadores pertencentes a uma distribuição normal com média zero e variância σ_g^2 (variância genética dividida pelo número de marcadores). Esses foram a chave do sucesso ao incluir toda a informação dos marcadores, além dos métodos cuja abordagem considerada é bayesiana, a qual Meuwissen, Hayes e Goddard (2001) propuseram os modelos chamados Bayes A e Bayes B, que relaxam a suposição de efeito marcadores iguais e variância comum e acomodam efeitos marcadores de diferentes tamanhos.

A partir do advento das inúmeras representações metodológicas, iniciaram-se buscas por redefinições dos problemas e conceitos da Seleção Genômica Ampla, o que produziu uma grande quantidade de modelos e métodos de GWS. Dos modelos de predição em GWS desenvolvidos, alguns são: (1) modelo de cumeeira; (2) modelos de abordagem Bayesiana; (3) modelo de regressão semi-paramétricos: Gianola, Fernando e Stella (2006) apresentam procedimentos de Regressão RKHS (*Reproducing Kernel Hilbert Space*); (4) métodos de aprendizagem de máquinas como *Random Forest* (JANNINK; LORENZ; IWATA, 2010).

Percebe-se que, mesmo com avanços consideráveis na tecnologia de marcadores, pode-se não superar determinadas características de um método em comparação ao outro. Um método pode ser mais informativo, interferindo e condicionando a eficiência da seleção. Um exemplo seria, pelo método RR-BLUP (*Ridge Regression-Best Linear Unbiased Prediction-RR-BLUP*), apesar da incorreta suposição de variância igual entre marcadores. Este método supera a Regressão *Stepwise*, porque estima simultaneamente efeito de todos os marcadores e evita vieses que acompanham a seleção (WHITTAKER; THOMPSON; DENHAM, 2000).

Existe uma variedade de métodos de seleção genômica ampla, os quais podem ser caracterizados em grupos, conforme o comportamento ou contribuição da variância dos SNPs (ver Figura 2.2).

Figura 2.2 – Classificação dos modelos de regressão em Sequenciamento Genômico



Fonte: Adaptado Desta e Ortiz (2014)

Apresenta-se, a seguir, a Tabela 2.1 com as características de cada método envolvido no estudo de seleção genômica ampla.

Tabela 2.1 – Principais características de alguns métodos preditivos utilizados em Seleção Genômica

| Métodos | Características | Referências |
|--------------|--|---|
| RR-BLUP | Assume que todos os marcadores têm variâncias iguais, mas com pequeno efeito não nulo. Aplica o <i>Shrinkage</i> homogêneo de preditores em direção a zero, mas permite que os marcadores tenham efeitos desiguais. | (MEUWISSEN; HAYES; GODDARD, 2001) (HEFFNER et al., 2011) |
| LASSO | Combina métodos de <i>Shrinkage</i> e de seleção de variável. RR-BLUP não usa seleção de variáveis, mas o LASSO supera quando há multicolinearidade entre os preditores. | (FRIEDMAN; HASTIE; TIBSHIRANI, 2010) (LI; SILLANPää, 2012) |
| EN | A regularização dupla, usando penalidades l_1 e l_2 , combina as características dessas normas para desafio de dados de alta dimensão. | (ZOU; HASTIE, 2005) |
| BRR | Induz o encolhimento homogêneo de todos os efeitos do marcador para zero e produz uma distribuição Gaussiana dos efeitos dos marcadores. Similar ao RR-BLUP, apresenta uma boa performance preditiva em arquiteturas infinitesimais. | (DE LOS CAMPOS et al., 2013) |
| BLASSO | Aplica-se ao método <i>Shrinkage</i> e à seleção de variáveis. Tem uma priori Exponencial sobre variância do marcador, resultando em uma distribuição Exponencial Dupla. A distribuição Exponencial Dupla tem uma maior densidade de massa em zero e cauda mais pesada, em comparação com distribuição Gaussiana. | (LI; SILLANPää, 2012) (DE LOS CAMPOS et al., 2009) |
| Bayes A | Utiliza distribuição qui-quadrado invertida escalonada χ^2 para variâncias dos marcadores, produzindo uma distribuição t escalonada para efeitos de marcador. Similar ao BLASSO, e em contraste com BRR, encolhe efeitos de marcador menores para zero e destaca valores maiores. Tem um pico mais alto de densidade de massa zero, em comparação com a distribuição Exponencial Dupla. | (MEUWISSEN; HAYES; GODDARD, 2001) (DE LOS CAMPOS et al., 2009) |
| Bayes B | Similar ao Bayes A, usa inversa χ^2 , resultando em uma distribuição t escalada. Diferentemente de Bayes A, utiliza métodos <i>Shrinkage</i> adicionando seleção de variáveis. Quando $\pi = 1$, então é similar a Bayes A. | (MEUWISSEN; HAYES; GODDARD, 2001) (HEFFNER; JANNINK; SORRELLS, 2011) |
| Bayes C | Aplica tanto o método <i>Shrinkage</i> , quanto os métodos de seleção de variáveis. Caracterizado por uma distribuição Gaussiana a priori, integra o efeito do marcador fora e seleciona com base na informação de variância. Bayes B e Bayes C consistem em ponto de massa zero, em suas priores. Bayes C supõe variância constante e π fixo. | (DE LOS CAMPOS et al., 2009) (DE LOS CAMPOS et al., 2013) |
| Bayes $C\pi$ | Uma variante modificada de Bayes B e Bayes C. Usado para atenuar as deficiências de Bayes C e Bayes B. Diferentemente de Bayes B, π não é fixo, mas estimado a partir dos dados. | (HABIER et al., 2011) |
| BSSV | Método semelhante ao Bayes B, porém não apresenta delta de Dirac como mistura a priori, mas, sim uma distribuição normal centrada no zero com variância pequena. Originalmente, utiliza uma regressão Bayesiana semelhante a uma suposição de efeito fixo, dado que se $\pi = 1$, a priori é assumida como gaussiana com variância alta. Pode ser entendido como um Bayes $B\pi$, se π é considerado desconhecido. | (YI; GEORGE; ALLISON, 2003) (DOS SANTOS et al., 2016) |
| GBLUP | Método RR-BLUP, com efeitos de marcas integrado para as componentes de variância. Utiliza informação da matriz de SNPs para a construção da matriz de parentesco. Caso específico dos modelos <i>Kernel</i> . | (VANRADEN, 2008) |
| RKHS | Com base na distância genética e função <i>kernel</i> , com um parâmetro regulador para Distribuição de efeitos QTL. Eficaz em situações as quais os efeitos não aditivos do gene aparecem. | (GIANOLA; KAAM, 2008) (DE LOS CAMPOS et al., 2010) |
| RF | Utiliza o modelo de regressão fixado nas observações de amostra <i>bootstrap</i> . Considera a média de todos os nós de árvore, para encontrar o melhor modelo de predição. Captura as interações entre marcadores. | (RUTKOSKI; HEFFNER; SORRELLS, 2011) (JANNINK; LORENZ; IWATA, 2010) (HOLLIDAY; WANG; AITKEN, 2012) |

EN:elastic net; RF: random forest; RKHS: Reproducing Kernel Hilbert Spaces

Fonte: Adaptado Desta e Ortiz (2014)

2.1.1 Modelos Usuais de Seleção Genômica

2.1.1.1 RR-BLUP

O método Clássico de Seleção Genômica, conhecido como Regressão Aleatória ou Regressão de Cumeira (*Ridge Regression - Best Linear Unbiased Prediction - RR-BLUP*), foi proposto pela primeira vez por Whittaker, Thompson e Denham (2000).

O uso do RR-BLUP permitiu grandes ganhos em estudos genéticos e nas metodologias tradicionais, que dependiam de informações da matriz de *pedrigree* para definir a estrutura de covariância (definida usando informações dos marcadores - matriz de parentesco genômico) (GONDRO et al., 2013).

Em propostas para empregar marcadores genômicos, Meuwissen, Hayes e Goddard (2001) examinam o método RR-BLUP, assumindo o modelo

$$\mathbf{y} = \boldsymbol{\mu} + \sum_k \mathbf{z}_k a_k + \mathbf{e} \quad (2.1)$$

em que \mathbf{y} é o vetor ($n \times 1$) de fenótipos, com n observações; $\boldsymbol{\mu}$ é a média geral, \mathbf{z}_k é o vetor coluna ($n \times 1$), de um total de k vetores da matriz \mathbf{Z} de ordem ($n \times 1$) que contém os códigos genotipados 0, 1 e 2, a_k é o efeito de cada SNP (num total de k do vetor de marcadores \mathbf{a}) e \mathbf{e} é o vetor ($n \times 1$) de efeitos residuais aleatórios. Os elementos de \mathbf{Z} , para cada coluna k , têm um quantidade $2p_k$ (com p_k é a frequência alélica do marcador no locus k , dada pela distribuição binomial, esperando que os alelos estejam em Equilíbrio de Hardy-Weinberg). Neste caso, o efeito dos SNPs são tratados como aleatórios e somados em todos os segmentos. A variância genética explicada pelos efeitos dos SNPs é dada por $\mathbf{Z}\mathbf{Z}'\sigma_a^2$ e a variância residual por $\mathbf{I}\sigma_e^2$. A matriz de variância-covariância entre as observações é

$$\text{Var}(\mathbf{y}) = \mathbf{V} = \mathbf{Z}\mathbf{Z}'\sigma_a^2 + \mathbf{I}\sigma_e^2 \quad (2.2)$$

Um critério encontrado, para verificar a qualidade do ajuste, minimizando o erro, é resolvendo as equações da soma de quadrados penalizada dada por

$$(\mathbf{y} - \mathbf{Z}\mathbf{a})'(\mathbf{y} - \mathbf{Z}\mathbf{a}) + \lambda \mathbf{a}'\mathbf{a}$$

Em que \mathbf{Za} é produto da matriz \mathbf{Z} pelo vetor de marcadores \mathbf{a} e a penalização é dada por $\mathbf{a}'\mathbf{a}$ e λ é um parâmetro de suavização, reconhecido como parâmetro de encolhimento (*shrinkage*) para efeitos aleatórios do SNPs. Os coeficientes de regressão são dados por

$$\hat{\mathbf{a}} = (\mathbf{Z}'\mathbf{Z} + \lambda\mathbf{I})^{-1}\mathbf{Z}'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \quad (2.3)$$

Assumindo-se variância igual para cada SNP, tem-se $\lambda = \frac{\sigma_e^2}{\sigma_a^2}$, determinado pelos estimadores do componente de variância. A organização do problema de predição, usando RR (*Ridge Regression*), está relacionada às predições que surgem sob um modelo linear misto. Na maioria dos modelos mistos que usam SNPs, a suposição é de que os efeitos do SNP são normalmente distribuídos com média zero e variância σ_a^2 . Os termos do resíduo também são normalmente distribuídos com variância σ_e^2 , os quais definem as matrizes $\mathbf{G} = \sigma_a^2\mathbf{I}$ e $\mathbf{R} = \sigma_e^2\mathbf{I}_n$.

2.1.1.2 Bayes A

Outro método usual em seleção genômica, denominado de Bayes A, (MEUWISSEN; HAYES; GODDARD, 2001) declaram que, é um método bayesiano de estrutura hierárquica para predição dos efeitos genéticos aditivos.

Pelo método Bayes A, a suposição inicial dos efeitos dos marcadores, para o k -ésimo marcador, com $i = 1, \dots, k$ é que tem distribuição normal univariada, com média zero e variância $\sigma_{a_i}^2$ específica para cada locus. A distribuição de variância dos marcadores é uma distribuição qui-quadrado invertida escalonada, com parâmetro escala s_a^2 e ν_a graus de liberdade.

$$a_i | \sigma_{a_i}^2 \sim N(0, \sigma_{a_i}^2) \quad (2.4)$$

$$\sigma_{a_i}^2 | \nu_a, s_a^2 \sim \chi^{-2}(\nu_a, s_a^2) \quad (2.5)$$

Desta forma, a distribuição marginal para a_i resume-se à distribuição t-Student univariada dada por

$$\begin{aligned} p(a_i | \nu_a, s_a^2) &= \int_0^\infty N(0, \sigma_{a_i}^2) p(\sigma_{a_i}^2 | \nu_a, s_a^2) d\sigma_{a_i}^2 \\ &\propto \int_0^\infty (\sigma_{a_i}^2)^{-((1+\nu_a+2)/2)} \exp\left[-\frac{a_i^2 + \nu_a s_a^2}{2\sigma_{a_i}^2}\right] d\sigma_{a_i}^2 \\ &\propto \left[1 + \frac{a_i^2}{\nu_a s_a^2}\right]^{-((\nu_a+1)/2)} \end{aligned} \quad (2.6)$$

com média nula e parâmetros escala s_a^2 e v_a graus de liberdade, assumindo que a variância residual é conhecida (GIANOLA et al., 2009) e (GIANOLA, 2013).

A modelagem do efeito dos marcadores, como resultado de uma distribuição t-Student univariada, utiliza amostrador de Gibbs e gera amostras da densidade a posteriori conjunta, ao final do processo MCMC.

2.1.1.3 Bayes B

O método Bayes B, que contém em si o método Bayes A, foi apresentado para estimar os valores genéticos genômicos, verificado por Meuwissen, Hayes e Goddard (2001). Amplamente utilizado nos estudos subsequentes de seleção genômica, destaca uma mistura de distribuições, com classificação a priori conjunta para a_i e $\sigma_{a_i}^2$ ($i = 1, \dots, k$), proposta como segue:

$$a_i | \sigma_{a_i}^2 \sim \begin{cases} \text{Ponto de massa para alguma constante } c, & \text{se } \sigma_{a_i}^2 = 0 \\ N(0, \sigma_a^2), & \text{se } \sigma_{a_i}^2 > 0 \end{cases} \quad (2.7)$$

$$\sigma_a^2 | \pi \sim \begin{cases} 0, & \text{com probabilidade } \pi \\ \chi^{-2}(v_a, s_a^2), & \text{com probabilidade } (1 - \pi) \end{cases} \quad (2.8)$$

Integrando-se a distribuição em $\sigma_{a_i}^2$, assume-se a forma:

$$a_i | v_a, s_a^2 \sim \begin{cases} 0, & \text{com probabilidade } \pi \\ t(0, v_a, s_a^2), & \text{com probabilidade } (1 - \pi) \end{cases} \quad (2.9)$$

Esta abordagem Bayesiana pode ser utilizada, em situações em que $n \ll k$, ou seja, em que o número de observações é muito menor que o número de marcadores. O efeito dos marcadores tem mistura de distribuições independentes, no qual cada uma tem centro de massa em zero, com probabilidade π e uma distribuição t-univariada com probabilidade $(1 - \pi)$ e média nula, parâmetro escala s_a^2 e v_a graus de liberdade. Esta mistura de distribuições é baseada na teoria de que muitos loci não têm efeito sobre as características e apenas poucos loci apresentam efeito e variância genética.

2.1.2 BSSV

O método BSSV (*Bayesian Stochastic Search Variable*) proposto por Yi, George e Allison (2003), baseia-se em um modelo bayesiano hierárquico, que incorpora uma mistura de distribuição, com capacidade de selecionar marcadores de grande efeito, em modelos com múltiplas marcas. Ao contrário do Bayes B, considera-se uma mistura de distribuição para o efeito dos marcadores como sendo uma variável aleatória de Bernoulli.

Para identificar marcadores que fortemente se ligam a genes com grandes efeitos e estimar a magnitude de seus efeitos, considera-se o modelo linear

$$y_i = \mu + \sum_{j=1}^k z_{ij}a_j + e_i \quad (2.10)$$

em que y_i é o vetor de fenótipos, μ é a média geral, z_{ij} é a marca j para o indivíduo i , com $i = 1, \dots, n$ e é definido como 0,5 ou $-0,5$, para os dois genótipos segregantes, na população de mapeamento, a_j é o efeito do marcador j com $j = 1, \dots, k$, e_i é o vetor de efeitos residuais aleatórios, considerando $e_i \sim N(0, \sigma_e^2)$.

Yi, George e Allison (2003) propuseram dois métodos para lidar com marcadores ausentes. Um primeiro método para substituir o genótipo z_{ij} ausente, por estatística multiponto, considerando esperança condicional de dado marcador e o segundo método de imputação de marcadores faltantes por amostragem, a partir da correspondente distribuição condicional completa.

Na seleção dos marcadores, pelo processo estocástico, considera-se um conjunto de variáveis indicadoras binárias, com $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_k)$ e γ_j sendo 1 ou 0, representando a presença ou a ausência da covariável j no modelo, respectivamente.

$$p(\boldsymbol{\gamma}) = \prod_{j=1}^k p(\gamma_j) \quad (2.11)$$

Quando nenhuma informação está disponível, a distribuição a priori uniforme é escolhida para cada γ_j , isto é, $p(\gamma_j = 0) = p(\gamma_j = 1) = 0,5$.

O efeito das marcas é dado pela distribuição condicional normal a priori em correspondência com o indicador γ_j

$$a_j | \gamma_j \sim (1 - \gamma_j) N(0, \tau_j^2) + \gamma_j N(0, c_j^2 \tau_j^2), j = 1, \dots, k \quad (2.12)$$

Os parâmetros τ_j^2 e $c_j^2\tau_j^2$ são escolhidos de tal forma que τ_j^2 é pequeno e $c_j^2\tau_j^2$ é grande. Consequentemente, se $\gamma_j = 0$, a magnitude do efeito do marcador é pequena e, então, a distribuição a priori força os parâmetros a zero. Caso $\gamma_j = 1$, a magnitude do efeito marcador será grande e uma estimativa diferente de zero deve ser incluída no modelo. A distribuição a posteriori reduzirá os valores dos dados a zero. Para tanto, uma distribuição conjunta a priori normal multivariada de \mathbf{a} condicionado a $\boldsymbol{\gamma}$, pode ser considerada como

$$\mathbf{a}|\boldsymbol{\gamma} \sim N_k(0, \mathbf{D}_\boldsymbol{\gamma}\mathbf{R}\mathbf{D}_\boldsymbol{\gamma}) \quad (2.13)$$

em que \mathbf{R} é a matriz de correlação usual designada como $\mathbf{R} = \mathbf{I}$ ou $\mathbf{R} \propto (\mathbf{Z}\mathbf{Z}')^{-1}$ e $\mathbf{D} = \text{diag}[a_1\tau_1, \dots, a_k\tau_k]$, com $a_i = 1$ se $\gamma_i = 0$ e $a_i = c_i$, se $\gamma_i = 1$. A distribuição a priori de $\boldsymbol{\mu}$ assume distribuição $N(\boldsymbol{\eta}, \tau^2)$, com média e variância especificadas. A priori para σ_a^2 é uma distribuição qui-quadrado invertida escalonada com hiperparâmetros ν_a e s_a^2

$$\sigma_a^2 \sim \chi^{-2}(\nu_a, s_a^2) \quad (2.14)$$

Segundo Yi, George e Allison (2003), com as especificações consideradas, os efeitos maiores aparecerão com mais frequência e pode-se usar o amostrador de Gibbs para gerar amostras.

A distribuição a posteriori condicional para o efeito de $\boldsymbol{\mu}$, procede conforme

$$\boldsymbol{\mu}|\mathbf{y}, \mathbf{z}, \mathbf{a}, \sigma_e^2 \sim N_k\left(\frac{\frac{\boldsymbol{\eta}}{\tau^2} + \sum_{i=1}^n (y_i - \sum_{j=1}^k z_{ij}a_j)/\sigma_e^2}{1/\tau^2 + n/\sigma_e^2}, \frac{1}{1/\tau^2 + n/\sigma_e^2}\right) \quad (2.15)$$

A distribuição a posteriori condicional, para o efeito de marcador \mathbf{a} , é normal multivariada $N_k\left([\mathbf{Z}\mathbf{Z}' + \sigma_e^2(\mathbf{D}_\boldsymbol{\gamma}\mathbf{R}\mathbf{D}_\boldsymbol{\gamma})^{-1}]^{-1}\mathbf{Z}'(\mathbf{y} - \boldsymbol{\mu}), [\mathbf{Z}\mathbf{Z}' + \sigma_e^2(\mathbf{D}_\boldsymbol{\gamma}\mathbf{R}\mathbf{D}_\boldsymbol{\gamma})^{-1}]^{-1}\right)$, de cálculo computacional oneroso, sendo amostrado de uma distribuição a posteriori condicional completa normal, com todos os efeitos dos marcadores, exceto o j , conforme comentado em Yi, George e Allison (2003).

A amostra de σ_e^2 segue uma distribuição a posteriori χ^{-2} condicional completa

$$\sigma_e^2|\mathbf{y}, \mathbf{z}, \boldsymbol{\mu}, \mathbf{a} \sim \chi^{-2}\left(\nu_0 + n, \frac{\nu_0\sigma_0^2 + \sum_{i=1}^n (y_i - \boldsymbol{\mu} - \sum_{j=1}^k z_{ij}a_j)^2}{\nu_0 + n}\right) \quad (2.16)$$

A amostra de γ_j como $p(\gamma_j|\mathbf{y}, \gamma_{(-j)}, \boldsymbol{\mu}, \mathbf{z}, \boldsymbol{\mu}, \mathbf{a}, \boldsymbol{\sigma}_e^2) = p(\gamma_j|\gamma_{(-j)}, \boldsymbol{\mu}, \mathbf{a}, \boldsymbol{\sigma}_e^2)$, segue a distribuição de Bernoulli tal qual

$$p(\gamma_j|\gamma_{(-j)}, \boldsymbol{\mu}, \mathbf{a}, \boldsymbol{\sigma}_e^2) = \frac{p(\mathbf{a}|\gamma_{(-j)}, \gamma_j = 1)p(\gamma_j = 1)}{p(\mathbf{a}|\gamma_{(-j)}, \gamma_j = 1)p(\gamma_j = 1) + p(\mathbf{a}|\gamma_{(-j)}, \gamma_j = 0)p(\gamma_j = 1)} \quad (2.17)$$

Em que $\gamma_{(-j)}$ denota os termos de $\boldsymbol{\gamma}$, exceto o γ_j . Modificações ao método BSSV foram propostas por dos Santos et al. (2016), modelando $\boldsymbol{\gamma}$ por uma distribuição Beta.

2.2 Associação Genômica Ampla (GWAS)

Nos últimos treze anos, muitas descobertas científicas e biológicas foram feitas por planejamento experimental de estudos de Associação Genômica Ampla (GWAS), como forma de identificar genes envolvidos em determinadas características de interesse, de acordo com estudos para análise de milhares de polimorfismos (variações no genoma). Dentre os estudos de GWAS, o trabalho introduzido por Yu et al. (2006) consistiu em identificar regiões causais que controlam caracteres quantitativos, visando ao escaneamento do genoma, o que significou um ganho aos estudos disponíveis, considerando os interesses de controlar estrutura de população e parentesco.

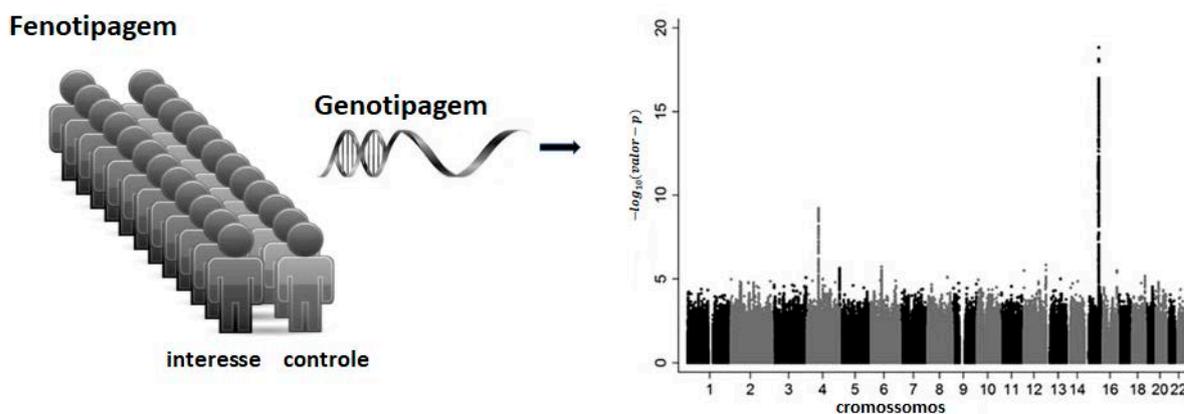
O objetivo da abordagem GWAS é fazer a varredura de marcadores SNPs, em todo o genoma, em busca de regiões causais associadas a uma determinada característica de interesse e, portanto responsáveis pela variação fenotípica. Outra consequência dessas associações, quando detectadas, é que podem estar atribuídas ao desequilíbrio de ligação (*linkage disequilibrium* LD) entre os marcadores, o que é explorável em GWAS.

As análises GWAS superaram as limitações de estudos de mapeamento de ligação para detecção de QTLs (*Quantitative Trait Loci*), pois a cobertura genômica utilizada em GWAS é ampla, não requer que as populações estejam em cruzamento, já que as associações genéticas detectadas são válidas para toda a população e não somente para o cruzamento. Além do mais, busca associar quantos e quais QTLs são responsáveis pela variação nas características fenotípicas, localizando a posição no genoma e estima os efeitos dos marcadores (ZHU et al., 2008).

De uma maneira simples, o fundamento inicial da análise GWAS segue por fazer uma regressão, para cada SNP genotipado, o número de cópias de alelo é regredido para a característica de interesse, entre todos os indivíduos. O resultado do valor-p para cada SNP é, então,

corrigido por testes múltiplos, e os efeitos significativos dos SNPs são identificados (LEISERSON et al., 2013). Os efeitos, como resultados do GWAS, são frequentemente exibidos em um gráfico de Manhattan (veja a Figura 2.3). Os valores-p são representados em ordem genômica por cromossomo e posição no cromossomo (eixo x). O valor no eixo y representa o $-\log_{10}(\text{valor} - p)$.

Figura 2.3 – Diagrama de identificação de variantes genéticas que distinguem populações de interesse e controle. A análise estatística GWAS calcula a associação entre um SNP para caso interesse/controle, referente ao $-\log_{10}(\text{valor} - p)$ para cada SNP



Fonte: Da autora (2019)

Segundo Visscher et al. (2017), o potencial de GWAS, para ter sucesso a uma característica de interesse, depende de (1) quantos loci afetam o caractere segregante na população, (2) arquitetura genética, (3) tamanho da amostra experimental, (4) painel de variantes genômicas que são usado nos GWAS e (5) quão heterogênea é a característica que está sendo estudada. Neste sentido, para a maioria das estratégias de GWAS, a compreensão tem relevância, para as principais questões científicas, especialmente, revelando a história demográfica da população em estudo. Ao incorporar a estrutura populacional e de parentesco aos métodos de GWAS, Yu et al. (2006) forneceram pertinente complemento aos modelos lineares mistos, o que compreende como a importância em se verificar o grau de relacionamento entre os indivíduos como propósito de não induzir a resultados falsos positivos.

Desde então, os métodos de GWAS, cuja finalidade é melhorar eficiência e utilidade na modelagem da estrutura populacional e parentesco entre indivíduos, foram baseados em modelos lineares mistos (Kang et al. (2008) e Zhou, Carbonetto e Stephens (2013)), ou ainda abordagem que inclui componentes principais como efeitos fixos a fim de modelar a estrutura de dependência (PRICE et al., 2006). No entanto as estatísticas que correspondem às diferenças no desempenho empírico entre a modelagem de componentes principais, com efeitos fixos versus

efeitos aleatórios, não foram exaustivamente examinadas, segundo Hoffman (2013), que realiza uma análise para formalizar a relação entre esses métodos utilizados em GWAS e elucidam as propriedades estatísticas, além de propor uma nova estatística.

Procedimentos propostos para estimar o conjunto de efeito de marcadores foram feitos e variam, conforme o objetivo do estudo, como nos estudos de Rakitsch et al. (2012) que permitiram diminuir o número de falsos positivos e maior resolução genômica, relatando a necessidade do modelo múltiplas marcas para GWAS como a proposta do MLM-LASSO (*Least Absolute Shrinkage and Selection Operator*), um modelo linear misto que permite mapeamento multiloci e correção para efeitos de confundimento, para tanto, modelando os efeitos genéticos complexos enquanto controla o parentesco. Loh et al. (2015) apresentam um método de associação de modelo misto, dito muito mais eficiente, o BOLT-LMM, o qual calcula estatísticas para testar a associação entre fenótipo e genótipos, usando modelo linear misto, modelando arquiteturas genéticas mais realistas e não infinitesimais, por meio de uma mistura de priori bayesiana para o efeito de marcador. Neste sentido, torna-se um atrativo o uso de modelos de seleção genômica aplicadas à GWAS.

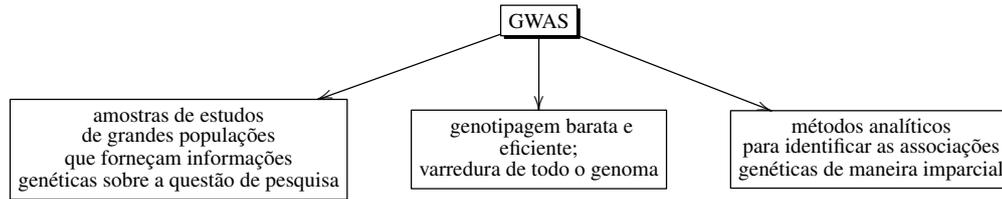
Outros trabalhos de associação (ZHANG et al., 2010) propõem modelos mistos adaptados aos estudos de associação genômica, com uma abordagem complementar de parâmetros da população previamente determinados (P3D), além dos trabalhos de Yang et al. (2010, 2011), tratando das estruturas da população por componentes principais da matriz construída à base de marcadores moleculares (conhecida como matriz G).

Destacam-se ainda, trabalhos recentes de Wang et al. (2016) e Wen et al. (2017). O primeiro desenvolve um algoritmo, para estimar variâncias dos marcadores e prever os efeitos desses marcadores, chamado de modelo RMLM (*random-SNP-effect mixed linear model - RMLM*) e o segundo implementa aplicação em múltiplos loci para a análise de GWAS com o método FASTmrEMMA.

2.2.1 Métodos Paramétricos para GWAS

Dentre os elementos essenciais requeridos para GWAS, segundo (CANTOR; LANGE; SINSHEIMER, 2010), os métodos analíticos e estatísticos podem ser empregados para identificar as associações genéticas (Figura 2.4).

Figura 2.4 – Representação dos elementos essenciais em GWAS



Fonte: Da autora (2019)

2.2.1.1 Análise de Associação Único Locus

A metodologia proposta por Yu et al. (2006) integra ao modelo estruturas populacional e de parentesco para GWAS, como alternativas para reduzir a detecção de falsos positivos ou a perda do poder estatístico pelo parentesco dentro da população. O modelo segue

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{S}\boldsymbol{\alpha} + \mathbf{Q}\mathbf{v} + \mathbf{Z}\boldsymbol{\mu} + \mathbf{e} \quad (2.18)$$

\mathbf{y} vetor de observações fenotípicas;

$\boldsymbol{\beta}$ vetor de efeitos fixos (outros efeitos diferentes dos marcadores e estrutura populacional);

$\boldsymbol{\alpha}$ vetor de efeito fixo do marcador SNP;

\mathbf{v} vetor de efeito fixo de estrutura da população

$\boldsymbol{\mu}$ vetor de efeitos aleatórios poligênicos desconhecidos (efeito poligênico *background*);

\mathbf{e} vetor aleatório de efeito residual;

\mathbf{Q} matriz relacionada à estrutura populacional;

\mathbf{X} , \mathbf{S} e \mathbf{Z} matrizes de incidência (compostas por zeros e uns), relacionando \mathbf{y} a $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$ e $\boldsymbol{\mu}$, respectivamente,

com distribuição dos efeitos $\boldsymbol{\mu} \sim N(0, \mathbf{K}\sigma_g^2)$ e $\mathbf{e} \sim N(0, \mathbf{R}\sigma_e^2)$, sendo σ_g^2 variância genética e σ_e^2 variância residual. A matriz $\mathbf{K}_{n \times n}$ é uma matriz de coeficientes de parentesco relativo define o grau de covariância entre um par de indivíduos; $\mathbf{R}_{n \times n}$ é uma matriz em que os elementos fora da diagonal são 0 e os elementos diagonais são os recíprocos do número de observações para

as quais cada ponto de dados fenotípico foi obtido. A proporção que envolve a razão entre variância genética e variância genética total é definida de herdabilidade (h^2)

$$h^2 = \frac{\sigma_g^2}{\sigma_g^2 + \sigma_e^2} \quad (2.19)$$

O efeito aleatório corrige a estrutura da população e o *background* genético simultaneamente. Modela a covariância genética entre os indivíduos da população, usando uma matriz de similaridade, denominada matriz de relacionamento genômico (*Genomic Relationship Matrix* - GRM). Durante a estimativa de efeito aleatório, cada marcador contribui com uma pequena fração, para a covariância dos indivíduos, assumindo que muitos loci com pequenos efeitos, ao longo de todo o genoma, estão contribuindo para o *background* genético (referido como fundo genético infinitesimal) (KLASEN et al., 2016).

A matriz \mathbf{K} é interpretada como matriz de parentesco ou matriz de relacionamento genômico entre indivíduos. Segundo Segura et al. (2012), as premissas diferentes levam a matrizes de parentesco diferentes, que podem ser usadas em modelos lineares mistos.

As estimativas (BLUE) de β , α e \mathbf{v} (efeitos fixos) e predições (BLUP) de μ (efeitos aleatórios) são obtidas, resolvendo-se equações de modelos lineares mistos (YU et al., 2006).

Outros métodos comuns para GWAS são versões baseados em um modelo linear misto de efeito fixo-SNP (MLM) e análise de marcador único, como a Análise Eficiente de Modelo Misto (*Efficient Mixed Model Analysis* - EMMA (KANG et al., 2008)). Esses métodos exigem a correção de Bonferroni, para vários testes, o que geralmente é muito conservador, quando o número de marcadores é extremamente grande (WANG et al., 2016).

2.2.1.1.1 Matriz de Parentesco Genômico

As covariâncias genéticas (parentesco genômico) entre os indivíduos são derivadas das probabilidades de que pares de genes compartilhados entre os indivíduos são idênticos por descendência (IBD- *identity by descent*), ou seja, o grau de parentesco entre os indivíduos é o dobro da probabilidade de, em um dado locus, um alelo retirado ao acaso de X e um alelo retirado ao acaso de Y serem idênticos por descendência. Tradicionalmente, a probabilidade de que dois alelos sejam IBD pode ser estimada utilizando informações de *pedigree* da população. Assim, utilizam-se da informação de *pedigree* para calcular a probabilidade de que dois indiví-

duos compartilhem o mesmo alelo, proveniente de um ancestral comum, montando a matriz de parentesco.

Powell e Goddard (2010) apresentaram conceitos e métodos, para calcular genes idênticos por descendência (IBD), que são usados para indicar a probabilidade de que dois alelos homólogos tenham sido herdados a partir de um ancestral comum.

A utilização de informações mais precisas sobre os genes IBD (*identity by descent*) e IBS (*identity by state*), compartilhadas por meio de ancestrais comuns, tornou-se possível pelos avanços tecnológicos e pelas diversas metodologias usadas, para calcular a matriz de parentesco, como observado em VanRaden (2008).

No método apresentado por VanRaden, em 2008, essencialmente, definiu-se sobre uma matriz de incidência \mathbf{W} , como a matriz de SNPs genotipadas, decodificada em termos numéricos 0, 1 e 2, que especifica quais alelos cada indivíduo herdou. A frequência do alelo no locus i é p_i . Essa matriz revela-se fundamental na estimação da matriz de parentesco genômico Yang et al. (2010). No caso, pode-se considerar o modelo linear misto ajustado para estimar os efeitos genéticos aditivos, usando informações fenotípicas e genotípicas

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e} \text{ com } \mathbf{g} = \mathbf{Z}\mathbf{u} \quad (2.20)$$

em que \mathbf{y} é o vetor $n \times 1$ de observações (fenótipos), $\boldsymbol{\beta}$ é o vetor de efeito fixo, \mathbf{u} é vetor de efeito genéticos aditivos tal qual $\mathbf{u} \sim N(0, \mathbf{I}\sigma_u^2)$, σ_u^2 é a variação do efeito causal; então, \mathbf{g} sendo o vetor de efeito genético total de $j = 1 \dots, n$ indivíduos com $g_j \sim N(0, \sigma_g^2 = k\sigma_u^2)$, σ_g^2 é a variância dos efeitos genéticos aditivos totais; \mathbf{e} é o vetor de resíduo. \mathbf{I} é a matriz identidade $n \times n$ e $\mathbf{e} \sim N(0, \mathbf{I}\sigma_e^2)$, \mathbf{X} e \mathbf{Z} são matrizes de incidência para os vetores $\boldsymbol{\beta}$ e \mathbf{g} , respectivamente. A matriz de variância-covariância de \mathbf{y} pode ser expressa por

$$\text{var}(\mathbf{y}) = \mathbf{Z}\mathbf{Z}'\sigma_u^2 + \mathbf{I}\sigma_e^2 = \frac{\mathbf{Z}\mathbf{Z}'}{k}\sigma_g^2 + \mathbf{I}\sigma_e^2 = \mathbf{G}\sigma_g^2 + \mathbf{I}\sigma_e^2 \quad (2.21)$$

em que \mathbf{G} é a matriz de relação genética entre os pares de indivíduos em locus causais. No entanto a matriz \mathbf{G} pode ser calculada, considerando uma amostra dos SNPs. Para tanto, o modelo 2.20, pode ser substituído por $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{W}\mathbf{a} + \mathbf{e}$, em que \mathbf{a} é um vetor de efeito dos marcadores. Após padronizar a matriz \mathbf{W} , a matriz de parentesco genômico é reconhecida como

$$\mathbf{A} = \frac{\mathbf{W}^*\mathbf{W}^{*'}}{2\sum p_i(1-p_i)} \quad (2.22)$$

e, neste caso, \mathbf{W}^* é a matriz de incidência dos marcadores corrigida pela média de cada locus p_i .

Outra interpretação para a matriz de parentesco é a proposta por Powell e Goddard (2010), como alternativa de considerar o relacionamento genético entre os indivíduos, considerando o conceito fundamental de identidade por descendência (alelos que são descendentes de um ancestral comum em uma população base), conhecida como matriz de parentesco unificada de Powell, na qual o coeficiente de parentesco entre indivíduos diferentes é dado por

$$\mathbf{A} = \begin{cases} \frac{1}{m} \sum_k \frac{(x_{ik}-2p_k)(x_{jk}-2p_k)}{2p_k(1-p_k)}, & \text{com } i \neq j \\ 1 + \frac{1}{m} \sum_k \frac{x_{ik}^2 - (1+2p_k)x_{jk} + 2p_k^2}{2p_k(1-p_k)}, & \text{com } i = j \end{cases} \quad (2.23)$$

em que \mathbf{x} é a soma de alelos de um dado locus de um indivíduo i ou j , em um locus k , com cada alelo codificado como 1 (favorável) ou 0 (desfavorável), p_k é a frequência de cada alelo, m é o número de SNPs na amostra.

Ao testar se um determinado locus tem um efeito sobre o fenótipo, em se tratando do uso do efeito aleatório, para corrigir a estrutura populacional e para um fundo genético infinitesimal, as hipóteses são alteradas. Em vez de testar se um determinado locus tem um efeito sobre o fenótipo, testa se um locus tem um efeito no fenótipo que não é explicado nem pela estrutura da população nem pelo fundo genético. Esta é uma restrição drástica à hipótese, que eventualmente leva a falhas na identificação de locus causais (KLASEN et al., 2016).

2.2.1.1.2 Testes Bonferroni e FDR

O teste de hipóteses sobre o coeficiente efeito de cada marcador a_k , com $k = 1, \dots, k$ é realizado, para verificar se há evidências de que esse coeficiente é estatisticamente diferente de zero, dado o nível de significância (α), ou seja, testar se rejeita a hipótese nula de que não existe relação linear entre o SNP e o fenótipo dados. Assim, as hipóteses nula e alternativa são

$$H_0 : a_k = 0$$

$$H_a : a_k \neq 0$$

O valor-p dá uma medida para a evidência contra a hipótese nula. Os testes estatísticos são geralmente ditos significativos e a hipótese nula é rejeitada, se o valor-p for igual ou menor que o nível de significância adotado a priori, em que, geralmente, $\alpha = 0,05$. Isto significa que

em 5% das vezes, a hipótese nula é rejeitada, quando, na verdade, é verdadeira e detecta-se um falso positivo. Como em GWAS, milhares de testes são conduzidos, cada um com sua própria probabilidade de falso positivo. A probabilidade cumulativa de encontrar um ou mais falsos positivos, em toda a análise GWAS é, portanto, muito maior, o que implica a necessidade de aplicação de alguma correção para testes múltiplos.

Uma das abordagens mais simples para corrigir testes múltiplos, é a correção de Bonferroni, que ajusta o valor de $\alpha = 0,05$, para $\alpha = 0,05/k$, em que k é o número de testes estatísticos realizados. Essa correção é a mais conservadora, pois assume que cada teste de associação do total de k é independente de todos os outros testes - uma suposição que geralmente é falsa em virtude do desequilíbrio de ligação entre os marcadores GWAS, porém é um método usual para fenótipos contínuos pela sua facilidade de aplicação e interpretação.

Uma alternativa para ajustar a taxa de falsos positivos (α) é determinar a taxa de falsa descoberta (FDR). A FDR é uma estimativa da proporção de resultados significativos (geralmente em $\alpha = 0,05$) que são falsos positivos. Sob a hipótese nula de que não há associações verdadeiras, em um conjunto de dados GWAS, os valores-p para testes de associação seguem uma distribuição uniforme. O ajuste normalmente utilizado, em análise genômica, utiliza a FDR desenvolvida por Benjamini e Hochberg (1995). Os procedimentos de FDR, essencialmente, são corretos para este número de falsas descobertas esperadas, fornecendo uma estimativa do número de resultados verdadeiros entre aqueles chamados significativos. Estas técnicas têm sido amplamente aplicadas ao GWAS e estendidas de várias maneiras, consistindo no procedimento:

1. Dada as m hipóteses, das m_0 verdadeiras, para cada hipótese H_i , calcula-se o valor-p correspondente a p_i ;
2. ordena-se de forma crescente de acordo com o valor-p

$$P_{(1)} \leq P_{(2)} \leq \dots \leq P_{(m)};$$

3. escolhe-se uma taxa de FDR arbitrária δ ;
4. o valor-p p_i são comparados ao valor crítico $\frac{i\delta}{m}$;
5. sendo $k = \max\{i : p_i < \frac{i\delta}{m}\}$ rejeita-se H_1, \dots, H_k , se k existir;
6. encerra-se a busca e adota-se $-\log_{10}(p_i)$ como valor crítico.

Esse procedimento permite controlar as taxas de falsas descobertas a partir de δ .

2.2.1.2 Análise de Associação Multiloci

Em contraste com a análise de associação de único locus, o conceito básico por trás dos métodos de análise multiloci é ajustar os efeitos de todos os SNPs, em geral, como efeitos aleatórios por um modelo linear misto. A análise multiloci, no entanto, não é tão simples quanto a realização de testes de único locus, uma vez que examinar combinações de pares de SNPs, entre milhões de SNPs genotipados em GWAS, ocasiona uma abordagem computacionalmente, muitas vezes, intratável, mesmo para algoritmos altamente eficientes (BUSH; MOORE, 2012). Uma proposta, para reduzir o tempo computacional necessário para GWAS, é o algoritmo P3D (parâmetros populacionais previamente definidos), adotado na etapa de varredura do genoma e fixa a razão da variância poligênico-residual (WANG et al., 2016).

As análises GWAS se adequaram a abordagens mais completas, incluindo todos os efeitos no modelo, que foi a proposta de Xu (2003), com avaliação de modelos múltiplos, utilizando regressão bayesiana e que passou a ser uma solução ideal para o problema de lidar com multicolinearidade. Várias outras abordagens multiloci foram declaradas, incluindo métodos de busca estocástica, com abordagem de máxima verossimilhança penalizada, inspirada em Bayesiana (HOGGART et al., 2008), LASSO Bayesiano (YI; XU, 2008) modelos lineares mistos LMM-Lasso (RAKITSCH et al., 2012).

As atuais evidências desenvolvidas, para metodologias GWAS multiloci, que visam superar problemas, envolvendo múltiplos testes e a correção Bonferroni, decorrem de seis métodos para GWAS multiloci, dos quais o modelo linear misto de efeito aleatório multiloci (*Multi-locus random effect mixed linear model* - MRMLM) é um método em que, no primeiro estágio, cada marcador é escaneado e exclui a maioria dos marcadores antecipadamente para que, no segundo estágio, todos os marcadores que são potencialmente associados à característica sejam incluídos no modelo multiloci - seus efeitos são estimados por Empírico Bayesiano-EM (WANG et al., 2016); o método FASTmrMLM, que reduz o tempo de execução do MRMLM em até 50% (TAMBA; ZHANG, 2018); o método chamado EMMA de efeito aleatório-SNP multiloci rápido - FASTmrEMMA (WEN et al., 2017), no qual o primeiro escolhe todos os possíveis nucleotídeos de característica quantitativa (QTNs) com *valores* $- p < 0,05$ e, em seguida, inclui em um modelo multiloci para detecção de QTN verdadeiro. Além do método ISIS EM-BLASSO de seleção e contração menos restritiva (BLASSO) utilizado, para estimar

todos os efeitos de SNP selecionados, para a detecção QTN (TAMBA; NI; ZHANG, 2017), pLARmEB e pKWmEB que estão implementados no pacote mrMLM do *software* R.

A representação seguinte trata-se do método MRMLM como um modelo de vários loci, incluindo marcadores selecionados do RMLM, e todos os efeitos do modelo são estimados por método Empírico Bayesiano-EM (WANG et al., 2016). A fundamentação teórica da análise Empírico Bayesiana EM pode ser encontrada em (XU, 2010).

A etapa inicial consiste em ajustar os efeitos de todos os SNPs como efeitos aleatórios por um modelo linear misto

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_k\gamma_k + \boldsymbol{\xi} + \mathbf{e} \quad (2.24)$$

no qual \mathbf{y} é o vetor $n \times 1$ de observações (fenótipos), $\boldsymbol{\beta}$ é o vetor de efeito fixo, \mathbf{Z}_k é um vetor de indicadores genotípicos para o k -ésimo SNP codificado como -1, 0, 1, para um dos dois homozigotos, o heterozigoto e o outro homozigoto, respectivamente; γ_k é o efeito do marcador k com uma distribuição normal assumida de média zero e variância ϕ_k^2 , $\boldsymbol{\xi} \sim N(0, \mathbf{K}\phi_k^2)$ é um vetor de efeitos poligênicos com uma distribuição normal multivariada de média zero e variância descrita ϕ_k^2 por uma estrutura de covariância \mathbf{K} , \mathbf{e} é um vetor de erros residuais com uma distribuição $N(0, \mathbf{I}\sigma^2)$. A esperança de $\mathbf{y} = \mathbf{X}\boldsymbol{\beta}$ e variância

$$\text{var}(\mathbf{y}) = [\mathbf{Z}_k\mathbf{Z}_k'\lambda_k + \mathbf{H}]\sigma^2$$

em que $\lambda_k = \frac{\phi_k^2}{\sigma^2}$ e $\mathbf{H} = \mathbf{K}\lambda + \mathbf{I}$, \mathbf{K} é a matriz de parentesco. Pelo método RMLM, a taxa da variância poligênicia λ é estimada apenas uma vez sob o modelo poligênico nulo antes do estágio da varredura do marcador. A razão da variância estimada $\hat{\lambda}$ é tratada como constante quando os marcadores são digitalizados. Como γ_k é tratado como aleatório, o algoritmo tratado por (WANG et al., 2016) de decomposição da matriz \mathbf{K}

A decomposição de $\mathbf{K} = \mathbf{U}\mathbf{D}\mathbf{U}'$, em que $\mathbf{D} = \text{diag}\{\delta_1, \dots, \delta_n\}$, é uma matriz diagonal de autovalores. Sendo $\mathbf{y}^* = \mathbf{U}'\mathbf{y}$, $\mathbf{X}^* = \mathbf{U}'\mathbf{X}$ e $\mathbf{Z}_k^* = \mathbf{U}'\mathbf{Z}_k$

$$\mathbf{y}^* = \mathbf{X}^*\boldsymbol{\beta} + \mathbf{Z}_k^*\gamma_k + \mathbf{U}'(\boldsymbol{\xi} + \mathbf{e}) \quad (2.25)$$

em que a esperança de \mathbf{y}^* é $E(\mathbf{y}^*) = \mathbf{X}^*\boldsymbol{\beta}$ e a matriz de variância e covariância de $\text{var}(\mathbf{y}^*) = [\mathbf{Z}_k^*\mathbf{Z}_k^*\lambda_k + \mathbf{R}]\sigma^2$, $\mathbf{R} = \mathbf{D}\hat{\lambda} + \mathbf{I}$ e $\mathbf{R}_k = \mathbf{Z}_k^*\mathbf{Z}_k^*\hat{\lambda}_k + \mathbf{R}$, Por meio da verossimilhança obtiveram-se as

estimativas por algoritmos de Newton para λ_k , e as estimativas de $\boldsymbol{\beta}$ e σ^2 são

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^{*\prime} \mathbf{R}_k \mathbf{X}^*)^{-1} \mathbf{X}^{*\prime} \mathbf{R}_k \mathbf{y}^* \quad (2.26)$$

$$\hat{\sigma}^2 = \frac{1}{n-q} (\mathbf{y}^* - \mathbf{X}^* \hat{\boldsymbol{\beta}})' \mathbf{R}_k^{-1} (\mathbf{y}^* - \mathbf{X}^* \hat{\boldsymbol{\beta}}) \quad (2.27)$$

A predição linear não viesada BLUP para γ_k é a esperança condicional de γ_k dado \mathbf{y}^* e variância condicional, cujas expressões seguem:

$$\hat{\gamma}_k = \lambda_k \mathbf{Z}_k^{*\prime} \mathbf{R}_k^{-1} \mathbf{y}^* - \lambda_k \mathbf{Z}_k^{*\prime} \mathbf{R}_k^{-1} \mathbf{X}^* [\mathbf{X}^{*\prime} \mathbf{R}_k^{-1} \mathbf{X}^*]^{-1} \mathbf{X}^{*\prime} \mathbf{R}_k^{-1} \mathbf{y}^* \quad (2.28)$$

$$\text{var}(\hat{\gamma}_k) = \lambda_k \sigma^2 - \lambda_k \mathbf{Z}_k^{*\prime} \mathbf{R}_k^{-1} \mathbf{Z}_k^* \lambda_k \sigma^2 \quad (2.29)$$

Para a abordagem RMLM de único locus, primeiro estima-se λ e depois fixado $\hat{\lambda}$ para estimar λ_k e realizar a varredura de todos os marcadores, testando $\lambda_k = 0$ para cada SNP. O teste de Wald foi implementado, com hipótese nula $H_0 : \lambda_k = 0$ ou $H_0 : \gamma_k = 0$,

$$W_k = \frac{\hat{\gamma}_k}{\text{var}(\hat{\gamma}_k)} \quad (2.30)$$

O valor $-p$ do teste de Wald é calculado por $p_k = 1 - \text{Pr}(\chi_1^2 < W_k)$, em que χ_1^2 é uma variável qui-quadrado com um grau de liberdade.

Em seguida, ajustam-se os marcadores pelo critério de Bonferroni modificado, usando um número eficaz de marcadores de modo que a limiar do *valor - p* é $\frac{0,005}{m_e}$, em que $m_e = \sum_{k=1}^m \frac{1}{\hat{\phi}_k^2} [\hat{\phi}_k^2 - \text{var}(\hat{\gamma}_k)]$

O modelo RMLM é considerado como etapa inicial de triagem para o novo modelo linear misto de efeito aleatório multiloci (MRMLM). No segundo estágio, após os marcadores selecionados, os marcadores restantes que passaram pela correção de Bonferroni modificada são usados para conduzir o teste da razão de verossimilhança e os outros são tratados como aleatórios. Para o valor crítico, considera-se o teste genético LOD maior que 1,5 trata o marcador como fixo. Os marcadores sobreviventes é que serão incluídos no modelo multiloci, como método Empírico Bayesiano EM (XU, 2010), que é uma abordagem de modelo aleatório a que cada efeito de marcador é atribuída uma distribuição empírica com variância.

2.2.1.3 Desenvolvimento teórico gerais sobre as diferenças entre GWS e GWAS

Apresentados os métodos de análise GWS e GWAS, seus princípios são derivados da mesma informação que decorre do uso de marcadores moleculares SNPs e apontadas diferenças metodológicas adotadas para cada análise, verifica-se que cada uma das formas de análise não consegue atender necessidades na totalidade de seleção e associação genômica. Uma principal diferença em evidência entre GWS e GWAS é que, nos estudos GWAS, requer testes de marcadores como já mencionados - Bonferroni, FDR, Wald - o que na seleção genômica não se aplica quanto à restrição da priori assumida ser de efeito nulo, restringindo as estimativas do efeito dos marcadores encolhidas a zero. Esse aspecto pode ser percebido ao considerar cada um dos métodos clássicos de seleção genômica. Inicialmente o método RR-BLUP, em que o parâmetro de encolhimento é dado por $\lambda = \frac{\sigma_e^2}{\sigma_a^2}$, com σ_e^2 variância residual e σ_a^2 variância genética para efeitos aleatórios dos SNPs. Ao observar o coeficiente de regressão

$$\begin{aligned}\hat{\mathbf{a}} &= (\mathbf{Z}'\mathbf{Z} + \lambda\mathbf{I})^{-1}\mathbf{Z}'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \\ &= (\mathbf{Z}'\mathbf{Z} + \frac{\sigma_e^2}{\sigma_a^2}\mathbf{I})^{-1}\mathbf{Z}'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})\end{aligned}\quad (2.31)$$

e assumindo-se a equação 2.31 quando $\sigma_a^2 \rightarrow 0$, resulta em $\hat{\mathbf{a}} \approx 0$.

Para o método Bayes A, verifica-se pelos resultados da predição estabelecidos por Gianola (2013), construídos pela iteração t ; o efeito do marcador é dado

$$\hat{a}^{[t+1]} = (\mathbf{Z}'\mathbf{Z} + \mathbf{W}_a^t)^{-1}\mathbf{Z}'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})\quad (2.32)$$

em que $\mathbf{W}_a^t = \text{Diag} \left\{ \frac{\sigma_e^2(1 + \frac{1}{v_a})}{s_a^2 \left(1 + \frac{a_j^{2t}}{s_a^2}\right)} \right\}$, com parâmetro escala s_a^2 e v_a graus de liberdade. Percebe-se que o elemento da matriz $\mathbf{W}_a = \{w_{jj}, a_j\}$ é semelhante ao inverso da matriz de variância-covariância genética, de tal modo que quanto maior w_{jj}, a_j , mais forte é o encolhimento para zero.

Considere ainda as situações: Se $v_a \rightarrow \infty$ resulta em $\lambda_j \rightarrow \frac{\sigma_e^2}{\sigma_a^2}$, ou seja, assemelha-se ao parâmetro de encolhimento do RR-BLUP. De outra forma, se v_a for um número finito, os marcadores, cujos efeitos estão próximos de zero, são fortemente reduzidos, em comparação aos que têm maiores valores absolutos. De acordo com os dois métodos, tanto RR-BLUP quanto

Bayes A não permitem informações adequadas sobre os efeitos de marcadores, o que, conforme se verifica pela própria distribuição a priori assumida, encolhe o efeito dos marcadores a zero.

Em seguida, o método Bayes B, verificado pela esperança posteriori do marcador

$$E_{BayesB}(a|\pi, s_a^2, v_a, \sigma_a^2, \mathbf{y}) = (1 - \pi)E_{BayesA}(a|\pi, s_a^2, v_a, \sigma_a^2, \mathbf{y})$$

indica que o encolhimento em direção a zero é mais forte que em Bayes A.

Já o método BSSV apresenta uma vantagem clara, que é o uso de duas distribuições prévias diferentes para os marcadores, que forcem as médias posteriores de marcadores insignificantes a serem próximas de zero e as distribuições posteriores de marcadores significativos a serem determinadas pelos dados.

Por estas circunstância observadas, em cada um dos métodos, percebe-se que as limitações matemáticas acabam sublinhando a impossibilidade de testes em GWS, o que torna difícil declarar identificação de regiões causais.

Outra questão que se verifica, para a distinção entre os estudos GWAS e GWS, é quanto à estrutura da população (matriz \mathbf{Q}). A aplicação de correção para a estrutura de população, não ocorre na seleção genômica, o que beneficiaria a predição é ignorado, enquanto se torna uma estrutura decisiva em GWAS, uma vez que previnem associações falso positivo que podem ser por causa de padrões estruturais e aumentam o poder pela aplicação de uma correção específica para essa estrutura (YANG et al., 2014). Entretanto Price et al. (2010) indicam que a contabilização da estrutura populacional é mais desafiadora quando a estrutura de parentesco também está presente, motivando o desenvolvimento de novos métodos.

O estudo de GWAS, ao verificar a relação entre cada um dos marcadores, usualmente trata a matriz \mathbf{G} , conhecida como matriz de relação genética entre os pares de indivíduos em locus causais, considerada como parte da estrutura de variância e covariância. No entanto a matriz \mathbf{G} que pode ser calculada, considerando uma amostra dos SNPs por diferentes maneiras e executada em GWAS, passa a apresentar inconvenientes entre incorporar estrutura de covariância total e modelar o marcador visto como fixo.

Para ilustrar, descrevemos a seguinte questão. Suponha dados consistentes de k marcadores e n indivíduos e teste o marcador j como efeito fixo no modelo misto apresentado por Yu et al. (2006). Escrevendo os genótipos dos marcadores como a matriz $n \times k$, tem-se que $\mathbf{Z} = [\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_j, \dots, \mathbf{z}_k]$, o vetor \mathbf{z}_j com $j = 1, \dots, k$, cujos elementos são códigos dos marcadores SNPs 0, 1 e 2 para o número do alelo em um indivíduo diploide. Assim, o modelo e a

matriz de parentesco \mathbf{G} podem ser apresentados como

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \underbrace{\mathbf{S}\boldsymbol{\alpha}_j}_{(*)} + \mathbf{Q}\mathbf{v} + \mathbf{Z}\boldsymbol{\mu} + \mathbf{e} \quad (2.33)$$

$$\mathbf{G} = \mathbf{Z}\mathbf{Z}'_{n \times n} = \mathbf{z}_1\mathbf{z}'_1 + \mathbf{z}_2\mathbf{z}'_2 + \dots + \underbrace{\mathbf{z}_j\mathbf{z}'_j}_{(**)} + \dots + \mathbf{z}_k\mathbf{z}'_k \quad (2.34)$$

em que (*) verifica-se a incidência sobre uma única marca j e \mathbf{G} representa a soma de cada matriz $n \times n$. Se o número k de marcadores for muito grande, a contribuição do marcador j apresentado em (**) é insignificante ao que é produzido pelos $k - 1$ marcadores (GIANOLA et al., 2016). No entanto a questão apresentada é contraditória, pois o marcador inicialmente suposto como fixo, também, comporta-se como efeito aleatório, verificando a redundância de efeito ao repetir informação e inconsistência de modelos.

Percebe-se que, para cada efeito a ser estimado do total de k marcadores em GWAS, a remoção do produto $\mathbf{z}_j\mathbf{z}'_j$, referente ao marcador j , passa a ser um processo computacional oneroso.

Outra questão também transcorre, ao tratar da estrutura populacional, em que a análise de componentes principais (PCA) é amplamente utilizada para detectar essa estrutura. Segundo Hoffman (2013), é comum aplicar a modelos lineares mistos ou incluir componentes principais como efeitos fixos, para modelar a estrutura de dependência e, geralmente, declarar a representação da estrutura de população extraíndo as componentes principais (PCA) da matriz genotípica. O problema, segundo Gianola et al. (2016) é que nos estudos de modelos mistos em GWAS por Componentes Principais, a matriz \mathbf{G} é utilizada sem modificações para extrair as componentes que representa a estrutura de populações.

A seguinte situação ilustra como a contribuição da matriz \mathbf{G} se estabelece de forma dupla na abordagem que inclui componentes principais. Considere a matriz de dados genotípicos \mathbf{G} pela decomposição do valor singular implícito à análise de componentes principais (PCA), com a forma

$$\mathbf{G} = \mathbf{U}\boldsymbol{\Delta}\mathbf{U}'$$

Em que $\mathbf{U} = [\mathbf{U}_1, \mathbf{U}_2, \dots, \mathbf{U}_n]$ matriz de autovetores de ordem $n \times n$ com os i primeiros componentes principais sendo as primeiras colunas de \mathbf{U} , $\boldsymbol{\Delta}$, $n \times n$ é a matriz diagonal que contém os

valores singulares λ_i , o modelo linear incluindo as componentes principais assume a forma

$$\begin{aligned}\mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \mathbf{S}\boldsymbol{\alpha}_j + \mathbf{U}\boldsymbol{\omega} + \mathbf{e}; \text{ com } \mathbf{g} = \mathbf{U}\boldsymbol{\omega} \text{ e } \boldsymbol{\omega} \sim N(0, \Delta\sigma_g^2) \\ \text{var}(\mathbf{y}) &= \mathbf{U}\Delta\mathbf{U}'\sigma_g^2 + \mathbf{I}\sigma_e^2 = \mathbf{G}\sigma_g^2 + \mathbf{I}\sigma_e^2\end{aligned}$$

Suponha que a regressão seja nos dois primeiros vetores tratados como fixos para explicar alguma estrutura. Assim, o modelo pode ser reescrito

$$\mathbf{y} = \{\mathbf{X}\boldsymbol{\beta} + \mathbf{S}\boldsymbol{\alpha}_j + \mathbf{U}_1\boldsymbol{\omega}_1 + \mathbf{U}_2\boldsymbol{\omega}_2\}_{\text{fixo}} + \{\mathbf{U}_{[-1-2]}\boldsymbol{\omega}_{[-1-2]} + \mathbf{e}\}_{\text{aleatório}}$$

em que $\mathbf{U}_{[-1-2]} = \mathbf{g}_{[-1-2]}$ é a matriz de parentesco genômico de ordem $n \times (n - 2)$, após remoção de seus dois primeiros vetores e $\boldsymbol{\omega}_{[-1-2]}$ é o vetor correspondente.

$$\text{var}(\mathbf{g}_{[-1-2]}) = \mathbf{G}_{[-1-2]} = \mathbf{G} - \begin{bmatrix} \mathbf{U}_1 & \mathbf{U}_2 \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} \mathbf{U}'_1 \\ \mathbf{U}'_2 \end{bmatrix}$$

Conforme esta representação, o que aparece na regressão tratado como fixo, deveria ser removido na contribuição da \mathbf{G} com base na decomposição de autovalores em GWAS, para evitar problemas de redundância. Pode-se perceber ao transformar a matriz de parentesco, que as componentes são todas incluídas essencialmente com efeito aleatório em vez de efeito fixo (HOFFMAN, 2013).

2.3 O Modelo Linear Misto

Um modelo estatístico, ao estabelecer que se pretende examinar a explicação de uma ou mais variáveis dependentes, por meio de efeitos fixos e ou aleatórios (além do erro experimental), comuns com outros efeitos fixos e média, é denominado de modelo linear misto.

Há várias abordagens para modelos lineares mistos, estudados os procedimentos de análise de variância (CASELLA; GEORGE, 1992), modelo linear clássico (HENDERSON, 1953) e também modelos hierárquicos (LEE; NELDER; PAWITAN, 2006), mas, segundo Gianola e Rosa (2015), as equações de modelo linear misto apareceram na literatura de teoria estatística tardia e com moderação, uma vez que na história das avaliações genéticas, desde 1988 os ci-

entistas que participavam de programas de melhoramento já utilizam como o modelo base da estatística genética e impulsionavam o uso do então chamado modelo animal.

A medição de n observações são modeladas na forma matricial pelo modelo linear misto dado por

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{a} + \mathbf{e} \quad (2.35)$$

em que

\mathbf{y} é um vetor ($n \times 1$) de observações, n é o número de observações;

\mathbf{X} é uma matriz ($n \times p$) de especificação dos efeitos fixos, p é o número de parâmetros associados aos efeitos fixos;

$\boldsymbol{\beta}$ é um vetor ($p \times 1$) dos parâmetros de efeitos fixos;

\mathbf{Z} é uma matriz ($n \times k$) de especificação dos efeitos aleatórios, k é o número de parâmetros associados aos efeitos aleatórios;

\mathbf{a} é um vetor ($k \times 1$) de efeitos aleatórios;

\mathbf{e} é um vetor ($n \times 1$) residual.

Assume-se que a distribuição conjunta dos vetores \mathbf{a} e \mathbf{e} é dada por

$$\begin{bmatrix} \mathbf{a} \\ \mathbf{e} \end{bmatrix} \sim N \left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{G} & \mathbf{0} \\ \mathbf{0} & \mathbf{R} \end{bmatrix} \right)$$

nos quais \mathbf{G} e \mathbf{R} são matrizes positivas definidas, sendo seus elementos, componentes de variância. Mais formalmente, a introdução de efeitos aleatórios ajuda a distinguir a média da população, tal qual $E(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}$ e a matriz de covariância de \mathbf{y} é:

$$\begin{aligned} \mathbf{V} &= \text{Var}(\mathbf{y}) = \text{Var}(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{a} + \mathbf{e}) \\ &= \text{Var}(\mathbf{X}\boldsymbol{\beta}) + \text{Var}(\mathbf{Z}\mathbf{a}) + \text{Var}(\mathbf{e}) \quad [\text{termos independentes}] \\ &= \text{Var}(\mathbf{Z}\mathbf{a}) + \text{Var}(\mathbf{e}) \quad [\text{variância dos efeitos fixos é zero}] \\ &= \mathbf{Z}\text{Var}(\mathbf{a})\mathbf{Z}' + \mathbf{R} \quad [\mathbf{Z} \text{ é constante}] \\ \text{cov}(\mathbf{y}) &= \mathbf{V} = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R} \end{aligned} \quad (2.36)$$

Finalmente, para o vetor \mathbf{y} , temos

$$\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta}, \mathbf{V})$$

2.3.1 Estimação por ML e REML

Uma questão importante em estatística é como lidar com parâmetro de interesse em um modelo. O objetivo em modelos lineares mistos é estimar os coeficiente de regressão de efeito fixo $\boldsymbol{\beta}$, bem como os parâmetros \mathbf{G} e \mathbf{R} . Porém a solução, para os componentes da variância \mathbf{G} e \mathbf{R} , não são de forma fechada. A variância residual \mathbf{R} depende da média, logo para cada valor da média, tem-se um correspondente em \mathbf{R} e $E(\hat{\mathbf{R}}) \neq \mathbf{R}$ e a média passa a ser um parâmetro perturbador ou *nuisance*.

Parâmetro perturbador é definido como qualquer parâmetro que não seja relevante ou de utilidade imediata, mas pode interferir na análise dos parâmetros que são de interesse. Uma vez que o parâmetro de perturbação pode ter impacto considerável sobre a inferência do parâmetro de interesse, a eliminação dos parâmetros perturbadores passaram a ser um problema complexo na teoria de verossimilhança e levaram a caminhos alternativos de abordagem.

Considerando estimar os parâmetros de variância, uma metodologia desenvolvida por Hartley e Rao trata por máxima verossimilhança (*maximum likelihood - ML estimation*), em 1967 e que depois passou a ser modificada por Patterson e Thompson em 1971, que é a estimação por máxima verossimilhança restrita ou residual (*Residual Maximum Likelihood - REML estimation*) (KNIGHT, 2008).

Ao ilustrarmos o método de estimativas por máxima verossimilhança, para os parâmetros do modelo linear misto, identificamos o sintoma do problema na estimação dos componentes de variância que são viciados.

Sob os pressupostos do modelo 2.35, pretendemos estimar os parâmetros desconhecidos $\boldsymbol{\beta}$ e σ_e^2 , em que se considere $\mathbf{V} = \mathbf{I}\sigma_e^2$, dos dados disponíveis (\mathbf{X} e \mathbf{y}). Assim, a distribuição dos dados completos é dada por

$$\begin{aligned} \mathbf{y} &\sim N(\mathbf{X}\boldsymbol{\beta}, \mathbf{I}\sigma_e^2) \\ p(\mathbf{y}|\boldsymbol{\beta}, \sigma_e^2) &= |\mathbf{I}\sigma_e^2|^{-\frac{1}{2}} \exp \left[-\frac{1}{2\sigma_e^2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right] \end{aligned}$$

A estimativa de máxima verossimilhança (ML) é obtida maximizando função log-verossimilhança correspondente

$$\ln p(\mathbf{y}|\boldsymbol{\beta}, \sigma_e^2) = -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln \sigma_e^2 - \frac{1}{2\sigma_e^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \quad (2.37)$$

e obtida como

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \quad (2.38)$$

$$\hat{\sigma}_e^2 = \frac{1}{n}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \quad (2.39)$$

em que n é o número de observações, $\hat{\boldsymbol{\beta}}$ é o valor mínimo, obtido por mínimos quadrados ordinários. Observe que o cálculo para $\hat{\sigma}_e^2$ depende do valor de $\hat{\boldsymbol{\beta}}$.

O viés de um estimador é a diferença entre a esperança do estimador $E(\hat{\sigma}_e^2)$ e o valor verdadeiro de σ_e^2 (SORENSEN; GIANOLA, 2007). Assim tem-se verificado que,

$$E(\hat{\sigma}_e^2) = \frac{n-1}{n}\sigma_e^2 < \sigma_e^2$$

portanto a estimativa do componente da variância é viesado.

Para se reduzir o viés, Patterson e Thompson (1971) procederam à partição da função de verossimilhança, introduziram a ideia de REML, tratando o efeito fixo do modelo como parâmetro perturbador. A intuição por trás do REML é maximizar uma probabilidade modificada que é livre de componente média, em oposição à probabilidade original, como em máxima verossimilhança (MV ou *maximum likelihood*-ML).

Assim, propõe-se o método de maximizar a verossimilhança de uma transformação ortogonal, a partir da matriz \mathbf{K} , tal que $\mathbf{K}\mathbf{X} = 0$, logo $\mathbf{K}\mathbf{X}\boldsymbol{\beta} = \mathbf{K}\mathbf{y} = 0$. Para tanto, a esperança resulta em $E(\mathbf{K}\mathbf{y}) = 0$ e variância e covariância $Var(\mathbf{K}\mathbf{y}) = \mathbf{K}\mathbf{V}\mathbf{K}'$. Assim, a verossimilhança obtida dessa transformação não depende dos efeitos fixos de $\boldsymbol{\beta}$. Em geral, $\mathbf{K} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. A distribuição dos dados é dada por

$$\mathbf{K}\mathbf{y} \sim N(0, \mathbf{K}\mathbf{V}\mathbf{K}')$$

Então, o logaritmo da função da máxima verossimilhança restrita resulta

$$\ln p(\mathbf{y}|\sigma_e^2) \propto -\frac{n}{2}\ln |\mathbf{K}\mathbf{V}\mathbf{K}'| - \frac{1}{2}[\mathbf{y}'\mathbf{K}'(\mathbf{K}\mathbf{V}\mathbf{K}')^{-1}\mathbf{K}\mathbf{y}] \quad (2.40)$$

Para encontrar estimativas dos parâmetros de variância por máxima verossimilhança, exige-se o processamento numérico dessas expressões (Equação 2.40), recorre-se a técnicas iterativas como Newton-Raphson, Score de Fisher, o algoritmo AI (uma variação do algoritmo *Newton-Raphson*), o algoritmo EM (*Expectation Maximization*) que também pode ser usado, a fim de facilitar o procedimento de cálculo da densidade a posteriori, quando os dados são incompletos e algoritmo PXEM (variação do algoritmo EM) são discutidos por Knight (2008).

A estimação pelo algoritmo EM, por exemplo, considerando o modelo linear misto, envolve cálculo da esperança condicional da função log-verossimilhança dos dados completos, dado os dados observados e envolve processos de maximização da esperança condicional quanto aos parâmetros de variância Knight (2008). Os efeitos aleatórios de \mathbf{a} podem ser dados perdidos ou observados e os dados completos são \mathbf{y} . Assim, a distribuição dos dados completos é dada por

$$\begin{aligned} \mathbf{y}|\mathbf{a} &\sim N(\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{a}, \mathbf{R}) & \text{com} & \quad p(\mathbf{y}|\mathbf{a}) = \frac{1}{(2\pi)^{\frac{n}{2}} \mathbf{R}^{\frac{1}{2}}} e^{-\frac{1}{2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{Z}\mathbf{a})'\mathbf{R}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{Z}\mathbf{a})} \\ \mathbf{a} &\sim N(\mathbf{0}, \mathbf{G}) & \text{com} & \quad p(\mathbf{a}) = \frac{1}{(2\pi)^{\frac{1}{2}} \mathbf{G}^{\frac{1}{2}}} e^{-\frac{1}{2}(\mathbf{a})'\mathbf{G}^{-1}(\mathbf{a})} \end{aligned} \quad (2.41)$$

Contudo as expressões obtidas pelo algoritmo EM nem sempre apresentam forma fechada, devem ser obtidos numericamente e podem ocasionar problemas na obtenção do efeito fixo do modelo.

Pelo fato de os parâmetros perturbadores terem impacto sobre o parâmetro de interesse e por criarem complicações envolvidas na teoria de verossimilhança, percebidos na desenvolvimento da solução do problema analítico, proposto tanto por ML quanto por REML, teriam como facilitador a marginalização dos efeitos fixos do modelo, o que não pode ser feito via integração. Um cenário interessante ao estudo e que preenche esta lacuna, ao propor alternativa de melhoria no modelo, passa pelo entendimento de análise bayesiana, cuja solução sucede do procedimento de integrar a posteriori conjunta, em relação ao parâmetro perturbador e trabalhar com a distribuição a posteriori marginal do parâmetro de interesse.

2.3.2 Solução Efeito Fixo e Aleatório

De posse dos valores paramétricos dos componentes de variância (\mathbf{G} e \mathbf{R}) ou de suas estimativas, o problema passa a ser estimar o vetor de efeitos fixos $\boldsymbol{\beta}$ e de prever o vetor de efeitos aleatórios \mathbf{a} .

Convencionalmente os estimadores de efeito fixo são chamados estimadores, e os estimadores de efeitos aleatórios são chamados preditores, os quais seguem as soluções BLUE e BLUP, respectivamente. Como afirmaram Gianola e Rosa (2015), o BLUP representou um paradigma de predição, logo que apresentado pela primeira vez em abordagens genéticas de criação de animais. É um vetor que não pode ser estimado, pois é aleatório e verifica-se que a função objetivo maximizada por Henderson é uma densidade a posteriori conjunta ou função penalizadora, comum em estatística Bayesiana. O método estabelecido por Henderson ficou conhecido como melhor preditor linear não viesado (*best linear unbiased prediction*-BLUP), resolvidos por chamadas equações de modelo misto, desenvolvidas por Henderson em 1948 (HENDERSON, 1984), comumente apresentadas de forma matricial

$$\begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}'\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \end{pmatrix} \begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{a}} \end{pmatrix} = \begin{pmatrix} \mathbf{X}'\mathbf{R}^{-1}\mathbf{y} \\ \mathbf{Z}'\mathbf{R}^{-1}\mathbf{y} \end{pmatrix}$$

A soluções para as equações são:

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= (\mathbf{X}' \underbrace{[\mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{Z}(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1})^{-1}\mathbf{Z}'\mathbf{R}^{-1}]}_{\mathbf{H}^{-1}} \mathbf{X})^{-1} \mathbf{X}' \underbrace{[\mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{Z}(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1})^{-1}\mathbf{Z}'\mathbf{R}^{-1}]}_{\mathbf{H}^{-1}} \mathbf{y} \\ &= (\mathbf{X}'\mathbf{H}^{-1}\mathbf{X})^{-1} \mathbf{X}'\mathbf{H}^{-1}\mathbf{y} \\ &= (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \quad \text{pois } \mathbf{H}^{-1} = [\mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{Z}(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1})^{-1}\mathbf{Z}'\mathbf{R}^{-1}]^{-1} = (\mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R})^{-1} = \mathbf{V}^{-1} \quad (2.42) \\ \hat{\mathbf{a}} &= (\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1})^{-1} \mathbf{Z}'\mathbf{R}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \\ &= \mathbf{G}\mathbf{Z}'\mathbf{H}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \end{aligned} \quad (2.43)$$

com BLUE de $\boldsymbol{\beta}$ e o BLUP de \mathbf{a} .

Quanto à solução apresentada, o sistema de equações de modelos lineares mistos pode ser obtido da matriz inversa $\mathbf{H}^{-1} = [\mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{Z}(\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1})^{-1}\mathbf{Z}'\mathbf{R}^{-1}]$, o que pode ser uma dificuldade em dados de melhoramento genéticos uma vez que a ordem da matriz (número de SNPs) é de centenas ou milhares.

Uma forma de se obter os valores dos efeitos de \mathbf{a} seria obter o preditor linear para o efeito de \mathbf{a} . Considere que a predição do resíduo pode ser descrita por um preditor linear, ou seja

$$\hat{\mathbf{e}} = \mathbf{y} - \mathbf{P}\mathbf{y} \quad (2.44)$$

em que $\mathbf{P} = \mathbf{X}(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}$. Pode-se recorrer a um preditor do modelo

$$\hat{\mathbf{a}} = \mathbf{b} - \mathbf{c}\mathbf{y} \quad (2.45)$$

A estimativa de $\hat{\mathbf{a}}$ é obtida pela minimização da expressão

$$\operatorname{argmin}_{\mathbf{b}, \mathbf{c}} E(\mathbf{a} - \mathbf{b} - \mathbf{c}\mathbf{y})^2 \quad (2.46)$$

em que $\mathbf{b} = \boldsymbol{\mu}_a - \mathbf{c}\boldsymbol{\mu}_y$ e $\mathbf{c} = \mathbf{V}_y^{-1}\mathbf{C}_{y,a}$, então, igualando o resultado a zero, obtemos o preditor linear da equação 2.46

$$\hat{\mathbf{a}} = \mathbf{C}_{y,a}\mathbf{V}_y^{-1}(\mathbf{y} - \boldsymbol{\mu}_y) = \mathbf{GZV}^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \quad (2.47)$$

em que $\boldsymbol{\mu}_a = 0$, $\boldsymbol{\mu}_y = \mathbf{X}\boldsymbol{\beta}$, $\mathbf{V} = \mathbf{ZGZ}' + \mathbf{R}$ e a matriz de covariância $\mathbf{C}_{y,a} = \mathbf{GZ}'$

Se \mathbf{y} e \mathbf{a} são variáveis aleatórias, pode-se obter a distribuição condicional $p(\mathbf{a}|\mathbf{y})$ e calcular a esperança $E(\mathbf{a}|\mathbf{y})$. Considerando que \mathbf{a}, \mathbf{y} seguem uma distribuição normal multivariada com a forma

$$p(\mathbf{a}, \mathbf{y}|\boldsymbol{\mu}, \mathbf{V}) \propto e^{-\frac{1}{2} \left\{ (\mathbf{a} - \boldsymbol{\mu}_a, \mathbf{y} - \boldsymbol{\mu}_y) \begin{bmatrix} \mathbf{V}_a & \mathbf{C}_{a,y} \\ \mathbf{C}_{y,a} & \mathbf{V}_y \end{bmatrix}^{-1} \begin{pmatrix} \mathbf{a} - \boldsymbol{\mu}_a \\ \mathbf{y} - \boldsymbol{\mu}_y \end{pmatrix} \right\}}$$

Em que $E(\mathbf{a}|\mathbf{y})$ corresponde ao preditor linear não viciado de \mathbf{a}

$$\begin{aligned} BLUP(\mathbf{a}) &= \hat{\mathbf{a}} = \mathbf{GZ}'(\mathbf{ZGZ}' + \mathbf{R})^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \\ var(\mathbf{a}) &= \mathbf{G} - \mathbf{GZ}'(\mathbf{ZGZ}' + \mathbf{R})^{-1}\mathbf{GZ} = (\mathbf{Z}'\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1})^{-1} \end{aligned}$$

2.4 Inferência Bayesiana

A inferência Bayesiana pressupõe a análise da probabilidade de ocorrência de um determinado assunto em uma população a partir de um subconjunto de elementos dessa população. Permite lidar com maior quantidade de informação que a inferência clássica (frequentista) em virtude da atualização da informação sobre os dados em que se observa. Basicamente a aná-

lise Bayesiana difere da análise frequentista pela inclusão da distribuição a priori e, assim, as atualizações sobre a probabilidade são mediante o cálculo da função distribuição a posteriori.

Supondo ter observado um dado y a partir de um modelo considerado, em que $p(y|\theta)$ é uma função de verossimilhança e definida uma distribuição para o parâmetro θ representada por uma priori informativa, descrita como densidade $p(\theta)$, obtém-se uma distribuição de θ , dado que as observações de y são conhecidas, com cálculo bayesiano pela distribuição $p(\theta|y)$, dado pelo teorema de Bayes:

$$p(\theta|y) = \frac{p(\theta, y)}{p(y)} = \frac{p(y|\theta)p(\theta)}{p(y)} = \frac{p(y|\theta)p(\theta)}{\int p(y|\theta)p(\theta)d\theta} \quad (2.48)$$

no qual $p(\theta|y)$ é a distribuição de θ , observa-se que se obteve y é chamada distribuição a posteriori. No segundo membro da igualdade, o numerador é uma densidade conjunta de θ e y , o denominador é uma função densidade marginal de y , conhecida como constante normalizadora e que não traz informação a respeito de θ ; $p(y|\theta) = L(\theta|y)$ é uma função de verossimilhança de cada um dos possíveis valores de θ ; $p(\theta)$ é a função densidade a priori. Assim, o teorema pode ser reescrito

$$p(\theta|y) \propto L(\theta|y)p(\theta)$$

Observa-se que, ao omitir o termo $\int p(y|\theta)p(\theta)d\theta$, a igualdade foi substituída por uma proporcionalidade.

As densidades marginais a posteriori de cada parâmetro ou de um conjunto k de parâmetros, ou seja, $\theta = (\theta_1, \dots, \theta_k)$, é dada por

$$p(\theta_i|y) = \int \dots \int p(\theta|y)d\theta_1, \dots, d\theta_{i-1}, d\theta_{i+1}, \dots, d\theta_k \quad (2.49)$$

em que $\theta_{-i} = (\theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_k)$ é o vetor de $k - 1$ parâmetros, com todas as coordenadas de θ , exceto a i -ésima. Para a resolução desta integral, na maioria da vezes, é impossível de obter solução analítica. Nestes casos, o amostrador de *Gibbs Sampling*, introduzido na estatística por Geman e Geman (1984), é uma alternativa do método Monte Carlo via Cadeia de Markov (MCMC). Outros métodos de simulação, como algoritmo de *Metropolis-Hastings* (CHIB; GREENBERG, 1995), podem ser utilizados no caso da distribuição a posteriori condicional completa não ter função densidade conhecida.

2.4.1 Uso da inferência Bayesiana para a marginalização de parâmetros perturbadores

Considerando o teorema de Bayes pela expressão 2.48 e seja o vetor de parâmetros dado por $\boldsymbol{\theta} = (\theta_1, \theta_2)$ (com rigor e sem perda de generalidade $k = 2$). Suponha que seja θ_1 o parâmetro de interesse, todos os demais parâmetros envolvidos (no caso θ_2) ocorrem como parâmetros perturbadores ou *nuisance*. Assim, neste caso, a densidade marginal correspondente pode ser reescrita como

$$\begin{aligned} p(\theta_1|y) &= \int p(\theta_1, \theta_2|y)d\theta_2 \\ &= \int p(\theta_1|\theta_2, y)p(\theta_2|y)d\theta_2 \end{aligned} \quad (2.50)$$

na qual

$p(\theta_1|\theta_2, y)$ é a distribuição a posteriori condicional de θ_1 dado θ_2 ,

A densidade a posteriori condicional de um parâmetro pode ser expressa como

$$p(\theta_1|\theta_2, y) = \frac{p(\theta_1, \theta_2|y)}{p(\theta_2|y)} \quad (2.51)$$

Nesta distribuição, o interesse é somente em relação a θ_1 , o que resulta no denominador simplesmente como parte da constante de integração. Assim, pode-se escrever

$$\begin{aligned} p(\theta_1|\theta_2, y) &\propto p(\theta_1, \theta_2|y) \\ &\propto L(\theta_1, \theta_2|y)p(\theta_1, \theta_2) \\ &\propto L(\theta_1, \theta_2|y)p(\theta_1|\theta_2) \\ &\propto L(\theta_1|\theta_2, y)p(\theta_1|\theta_2) \end{aligned} \quad (2.52)$$

$L(\theta_1|\theta_2, y)$ é uma função de verossimilhança que trata θ_2 como constante. O desenvolvimento implica que uma distribuição a posteriori condicional pode (muitas vezes) ser identificada pela inspeção da densidade a posteriori conjunta, restringindo as partes que variam com os parâmetros de interesse, tratando o restante como conhecido. Para facilitar as tarefas de identificação das distribuições a posteriori, existem métodos numéricos, específicos e apropriados a cada situação. Os métodos de simulação Monte Carlo é um deles, bem como algoritmos de simulação Monte Carlo via Cadeias de Markov.

Uma situação exemplo, assumindo-se as características fenotípicas consideradas pelo modelo linear misto, é controlada pelo efeito de marcadores do tipo SNPs; a marginalização dos efeitos dos marcadores SNPs foi apresentada no Apêndice A.

2.4.2 Modelagem Hierárquica

Em situações de modelagem com muitos parâmetros, dispõe-se de distribuição a priori, de forma hierárquica. Considerando o modelo, as observações seguem distribuições condicionais em parâmetros. Os parâmetros, por sua vez, têm distribuições condicionais em parâmetros adicionais chamados hiperparâmetros. Ponderando q a função distribuição dos dados

$$y \sim f(y|\theta)$$

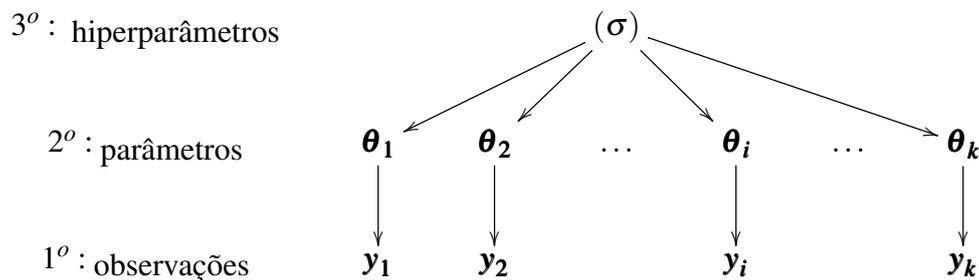
ao vetor θ é atribuída uma distribuição com hiperparâmetros desconhecidos σ

$$\theta \sim h_1(\theta|\sigma)$$

e ao hiperparâmetro σ segue a distribuição

$$\sigma \sim h_2(\sigma)$$

O esquema, a seguir, exemplifica os hiperparâmetros de modelos hierárquicos, para um conjunto de parâmetros $\theta = (\theta_1, \dots, \theta_k)$, considerando três níveis:



Assim, a estratégia de análise é construída sobre um modelo hierárquico de três níveis, em que θ é uma amostra aleatória de $h_1(\theta|\sigma)$ e o vetor σ de hiperparâmetro desconhecido é atribuído como conhecido no segundo nível. Os modelos hierárquicos bayesianos são mais realistas e, possivelmente, descrevem melhor o sistema biológico em muitos estudos, com metodologia que

permite ajustar, de forma maleável, fatores genéticos e ambientais (SORENSEN; GIANOLA, 2007).

2.4.3 Técnica Computacional

Os métodos MCMC (do inglês *Markov Chain Monte Carlo*) produzem aproximações empíricas para distribuições de probabilidade. São comumente usados em inferência estatísticas, com abordagem bayesiana, na qual os parâmetros desconhecidos são tratados como aleatórios e representados com distribuições de probabilidade (GELMAN et al., 2014). Dentre os métodos, destacam-se os algoritmos *Metropolis-Hastings* e o amostrador de *Gibbs*.

O amostrador de *Gibbs* é uma forma de amostragem a partir da distribuição de probabilidade e um caso particular do *Metropolis-Hastings*. No entanto, ao contrário do algoritmo *Metropolis-Hastings*, que aceita todas as propostas de amostragem (distribuições não identificáveis), na amostragem por *Gibbs*, é possível amostrar diretamente das distribuições a posteriori condicionais dos parâmetros, em vez de marginalizar integrando-se sobre uma distribuição conjunta (GELMAN et al., 2014). Demonstra-se que uma distribuição conjunta pode ser definida pelo conjunto completo de tais distribuições condicionais completas (BESAG, 1974).

Considere que o vetor de parâmetros $\boldsymbol{\theta}$ seja um vetor d -dimensional, $\boldsymbol{\theta} = (\theta_1, \dots, \theta_d) \in \Theta \subseteq \mathbb{R}^d$. O amostrador de *Gibbs* consiste em selecionar um componente aleatório i de $\boldsymbol{\theta}$ e substituí-lo por uma amostra de $p(\theta_i | \mathbf{y}, \boldsymbol{\theta}_{-i})$, ou seja, uma amostra determinada pelas distribuições condicionais completa, na qual $\boldsymbol{\theta}_{-i} = (\theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_d)$ é o vetor de $d - 1$ parâmetros, com todas as coordenadas de $\boldsymbol{\theta}$, exceto a i -ésima.

A descrição do algoritmo de *Gibbs* passa pelos seguintes passos:

- (1) Inicialize atribuindo valores aos parâmetros $\boldsymbol{\theta}^{(t)} = (\theta_1^{(t)}, \theta_2^{(t)}, \dots, \theta_d^{(t)})$ para $t = 0$.
- (2) Para $d = 1, \dots, d$, os valores para o parâmetro θ_i são obtidos, a partir da distribuição condicional completa, dados os valores atuais de todos os outros parâmetros do modelo e os dados observados. Ou seja, para cada parâmetro θ_i , obtém-se o $(t + 1)$ -ésimo valor da cadeia a partir de $p(\theta_i | \mathbf{y}, \theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_d)$. Um ciclo é dado, sequencialmente,

mostrando valores de

$$\begin{aligned}\theta_1^{(t+1)} &\sim p(\theta_1|\mathbf{y}, \theta_2^{(t+1)}, \dots, \theta_d^{(t+1)}) \\ \theta_2^{(t+1)} &\sim p(\theta_2|\mathbf{y}, \theta_1^{(t+1)}, \theta_3^{(t+1)}, \dots, \theta_d^{(t+1)}) \\ &\vdots \\ \theta_d^{(t+1)} &\sim p(\theta_d|\mathbf{y}, \theta_1^{(t+1)}, \theta_3^{(t+1)}, \dots, \theta_{d-1}^{(t+1)})\end{aligned}$$

(3) Repita a etapa (2) por um número grande de t iterações.

A sequência $\boldsymbol{\theta}^{(t)}$, $t = 1, 2, \dots, t$ é uma efetuação de uma cadeia de Markov. Depois de muitas iterações (quando $t \rightarrow \infty$), a amostra convergirá para, aproximadamente, a distribuição conjunta de todas as variáveis (BERNARDO; SMITH, 1994).

As amostras do início da cadeia podem não representar, com precisão, a distribuição desejada e são, geralmente, descartadas (o período de *burn-in*). Por outro lado, em cada iteração, o amostrador de Gibbs gera a amostragem de cada distribuição condicional e utiliza o novo valor para amostrar as outras distribuições condicionais. Neste caso, se houver uma autocorrelação elevada entre os vários estados da cadeia, deve-se escolher amostras espaçadas, de modo que garantam a não correlação (salto denominado *jump*).

Frequentemente, o valor esperado de qualquer variável pode ser aproximado pela média de todas as amostras e passa a descrever o conhecimento sobre o parâmetro de interesse. A simulação da distribuição a posteriori pelo amostrador de Gibbs envolve a simulação das condicionais completas. Porém, quando a simulação não é acessível, podem-se usar outras técnicas para simular as distribuições condicionais. Dentre as técnicas e detalhes sobre algoritmos, há vários trabalhos como (BERNARDO; SMITH, 1994), (GELMAN et al., 2014) e (SORENSEN; GIANOLA, 2007).

2.5 Formulação Modelo Múltiplas Marcas para GWAS

Consideramos um modelo linear misto e discutiremos a obtenção da distribuição a posteriori condicional dos marcadores por métodos de eliminação de parâmetros perturbadores (*nuisance*) e apontaremos informações retratadas pela matriz de parentesco.

Primeiro utilizou-se um modelo misto para correções de observação, de acordo com os efeitos: estrutura da população, genótipos e resíduos. O modelo misto adotado foi:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{a} + \mathbf{e} \quad (2.53)$$

Com o vetor $\mathbf{a}_{k \times 1} = (a_1, a_2, \dots, a_k)'$ intitulado pela conjuntura genética como vetor de efeitos aleatórios dos marcadores e $\mathbf{Z}_{n \times k} = \begin{bmatrix} \mathbf{z}_1 & \mathbf{z}_2 & \dots & \mathbf{z}_k \end{bmatrix}$ uma matriz de incidência dos efeito marcadores, podemos reescrever (2.53) :

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{z}_1 a_1 + \mathbf{z}_2 a_2 + \dots + \mathbf{z}_k a_k + \mathbf{e} \quad (2.54)$$

A partir do conhecimento do modelo de Henderson Equação 2.54, a dissecação analítica é estruturada e realizada pela Abordagem Hierárquica Bayesiana, na qual a distribuição a posteriori dos parâmetros é proporcional ao produto da distribuição condicional dos dados amostrais (função de verossimilhança) pelas distribuições a priori para os parâmetros de interesse. A ideia de hierarquia é dividir a especificação da distribuição a priori em estágios, facilitando a precisão desta abordagem.

A função densidade de probabilidade da distribuição condicional do vetor de observações \mathbf{y} é dada

$$\mathbf{y}|\mathbf{a} = \mathbf{y} | (\boldsymbol{\beta}, a_1, \dots, a_k) \sim N(\mathbf{X}\boldsymbol{\beta} + \mathbf{z}_1 a_1 + \dots + \mathbf{z}_k a_k, \mathbf{I}\sigma_e^2)$$

Ou ainda:

$$p(\mathbf{y}|\mathbf{a}) = p(\mathbf{y}|\boldsymbol{\beta}, a_1, \dots, a_k) \approx e^{-\frac{1}{2} \frac{(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{z}_1 a_1 - \dots - \mathbf{z}_k a_k)' (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{z}_1 a_1 - \dots - \mathbf{z}_k a_k)}{\sigma_e^2}} \quad (2.55)$$

que fornece a verossimilhança de cada um dos possíveis valores de \mathbf{y} .

No primeiro estágio da hierarquia, foram assumidas as distribuições a priori para cada efeito de SNP (parâmetro) do modelo:

$$\begin{aligned} a_1 &\sim N(0, \sigma_{a_1}^2) \\ &\vdots \\ a_k &\sim N(0, \sigma_{a_k}^2) \end{aligned}$$

Vale ressaltar que os efeitos dos SNPs são considerados independentes, mas cada SNP apresenta uma variância diferente. Se $\sigma_{a_1}^2 = \dots = \sigma_{a_k}^2$, então, a priori corresponde à classificação para RR-BLUP.

Para o modelo (2.54), um dos interesses principais reside na estimação dos marcadores; assim, considerando a distribuição a posteriori condicional, para os marcadores, escrevendo $\mathbf{R} = \mathbf{I}\sigma_e^2$, temos:

$$\begin{aligned} p(\mathbf{a}|\boldsymbol{\beta}, \sigma_a^2, \sigma_e^2, \mathbf{y}) &\propto e^{-\frac{1}{2} \frac{(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{z}_1 a_1 - \dots - \mathbf{z}_k a_k)'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{z}_1 a_1 - \dots - \mathbf{z}_k a_k)}{\sigma_e^2}} p(a_1) \cdots p(a_k) \\ &\propto e^{-\frac{1}{2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{z}_1 a_1 - \dots - \mathbf{z}_k a_k)'\mathbf{R}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{z}_1 a_1 - \dots - \mathbf{z}_k a_k)} p(a_1) \cdots p(a_k) \end{aligned} \quad (2.56)$$

Declarando para o vetor de parâmetros $\mathbf{a} = (a_1, \dots, a_k)$, a_1 é o primeiro parâmetro de interesse e $\mathbf{a}_{-i} = (a_2, \dots, a_k)$ o vetor de parâmetros perturbadores ou *nuisance*; o objetivo é eliminar os parâmetros perturbadores, supondo (a_1, \dots, a_k) , variáveis aleatórias independentes, com a mesma distribuição normal $(0, \sigma_{a_j}^2)$, com $j \in \{1, 2, \dots, k\}$.

Com rigor e sem perda de generalidade, passaremos a assumir $k = 2$. Supondo a_1 o parâmetro de interesse e a_2 o parâmetro perturbador, para reduzir a demonstração em poucos casos, com propósito de induzir o desenvolvimento analítico de forma instrutiva, didática e ilustrativa, desse modo, a função distribuição condicional a posteriori integrada é obtida por integrais múltiplas da função distribuição (2.56), varrendo o domínio de integração com os $(k - 1)$ parâmetros perturbadores (conceitue marcador a_2). Considerando o primeiro marcador a_1 como de interesse, demonstra-se:

$$\begin{aligned} p(a_1|\boldsymbol{\beta}, a_2, \sigma_a^2, \sigma_e^2, \mathbf{y}) &\propto \int e^{-\frac{1}{2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{z}_1 a_1 - \mathbf{z}_2 a_2)'\mathbf{R}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{z}_1 a_1 - \mathbf{z}_2 a_2)} p(a_1)p(a_2) da_2 \\ &\propto p(a_1) \int e^{-\frac{1}{2}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{z}_1 a_1 - \mathbf{z}_2 a_2)'\mathbf{R}^{-1}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{z}_1 a_1 - \mathbf{z}_2 a_2)} e^{-\frac{a_2(\sigma_{a_2}^2)^{-1} a_2}{2}} da_2 \end{aligned}$$

denotando por $\tilde{\mathbf{y}} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{z}_1 a_1$ e, na sequência, note que escrevemos $[\mathbf{z}'_2 \mathbf{R}^{-1} \mathbf{z}_2 + (\boldsymbol{\sigma}_{a_2}^2)^{-1}] = \boldsymbol{\Omega}^{-1}$, então, segue:

$$\begin{aligned}
p(a_1 | \boldsymbol{\beta}, a_2, \boldsymbol{\sigma}_a^2, \boldsymbol{\sigma}_e^2, \mathbf{y}) &\propto p(a_1) \int e^{-\frac{1}{2} [(\tilde{\mathbf{y}} - \mathbf{z}_2 a_2)' \mathbf{R}^{-1} (\tilde{\mathbf{y}} - \mathbf{z}_2 a_2) + a_2' (\boldsymbol{\sigma}_{a_2}^2)^{-1} a_2]} da_2 \\
&\propto p(a_1) \int e^{-\frac{1}{2} [\tilde{\mathbf{y}}' \mathbf{R}^{-1} \tilde{\mathbf{y}} - 2a_2' \mathbf{z}'_2 \mathbf{R}^{-1} \tilde{\mathbf{y}} + a_2' \mathbf{z}'_2 \mathbf{R}^{-1} \mathbf{z}_2 a_2 + a_2' (\boldsymbol{\sigma}_{a_2}^2)^{-1} a_2]} da_2 \\
&\propto p(a_1) \int e^{-\frac{1}{2} [\tilde{\mathbf{y}}' \mathbf{R}^{-1} \tilde{\mathbf{y}} - 2a_2' \mathbf{z}'_2 \mathbf{R}^{-1} \tilde{\mathbf{y}} + a_2' [\mathbf{z}'_2 \mathbf{R}^{-1} \mathbf{z}_2 + (\boldsymbol{\sigma}_{a_2}^2)^{-1}] a_2]} da_2 \\
&\propto p(a_1) \int e^{-\frac{1}{2} [\tilde{\mathbf{y}}' \mathbf{R}^{-1} \tilde{\mathbf{y}} - 2a_2' \mathbf{z}'_2 \mathbf{R}^{-1} \tilde{\mathbf{y}} + a_2' \boldsymbol{\Omega}^{-1} a_2]} da_2 \\
&\propto p(a_1) \int e^{-\frac{1}{2} [\tilde{\mathbf{y}}' \mathbf{R}^{-1} \tilde{\mathbf{y}} - 2a_2' \mathbf{z}'_2 \mathbf{R}^{-1} \tilde{\mathbf{y}} + a_2' \boldsymbol{\Omega}^{-1} a_2]} da_2 \\
&\propto p(a_1) \int e^{-\frac{1}{2} [\tilde{\mathbf{y}}' \mathbf{R}^{-1} \tilde{\mathbf{y}} - 2a_2' \boldsymbol{\Omega}^{-1} \boldsymbol{\Omega} \mathbf{z}'_2 \mathbf{V}^{-1} \tilde{\mathbf{y}} + a_2' \boldsymbol{\Omega}^{-1} a_2 + \tilde{\mathbf{y}}' \mathbf{R}^{-1} \mathbf{z}_2 \boldsymbol{\Omega} \boldsymbol{\Omega}^{-1} \boldsymbol{\Omega} \mathbf{z}'_2 \mathbf{R}^{-1} \tilde{\mathbf{y}} - \tilde{\mathbf{y}}' \mathbf{V}^{-1} \mathbf{z}_2 \boldsymbol{\Omega} \boldsymbol{\Omega}^{-1} \boldsymbol{\Omega} \mathbf{z}'_2 \mathbf{R}^{-1} \tilde{\mathbf{y}}]} da_2 \\
&\propto p(a_1) \int e^{-\frac{1}{2} [\tilde{\mathbf{y}}' \mathbf{R}^{-1} \tilde{\mathbf{y}} - \tilde{\mathbf{y}}' \mathbf{R}^{-1} \mathbf{z}_2 \boldsymbol{\Omega} \boldsymbol{\Omega}^{-1} \boldsymbol{\Omega} \mathbf{z}'_2 \mathbf{R}^{-1} \tilde{\mathbf{y}} + (a_2 - \boldsymbol{\Omega} \mathbf{z}'_2 \mathbf{R}^{-1} \tilde{\mathbf{y}})' \boldsymbol{\Omega}^{-1} (a_2 - \boldsymbol{\Omega} \mathbf{z}'_2 \mathbf{R}^{-1} \tilde{\mathbf{y}})]} da_2 \\
&\propto p(a_1) \int e^{-\frac{1}{2} [\tilde{\mathbf{y}}' \mathbf{R}^{-1} \tilde{\mathbf{y}} - \tilde{\mathbf{y}}' \mathbf{R}^{-1} \mathbf{z}_2 \boldsymbol{\Omega} \boldsymbol{\Omega}^{-1} \boldsymbol{\Omega} \mathbf{z}'_2 \mathbf{R}^{-1} \tilde{\mathbf{y}}]} e^{-\frac{1}{2} [(a_2 - \boldsymbol{\Omega} \mathbf{z}'_2 \mathbf{R}^{-1} \tilde{\mathbf{y}})' \boldsymbol{\Omega}^{-1} (a_2 - \boldsymbol{\Omega} \mathbf{z}'_2 \mathbf{R}^{-1} \tilde{\mathbf{y}})]} da_2 \quad (2.57)
\end{aligned}$$

verifica-se que $e^{-\frac{1}{2} [\tilde{\mathbf{y}}' \mathbf{R}^{-1} \tilde{\mathbf{y}} - \tilde{\mathbf{y}}' \mathbf{R}^{-1} \mathbf{z}_2 \boldsymbol{\Omega} \boldsymbol{\Omega}^{-1} \boldsymbol{\Omega} \mathbf{z}'_2 \mathbf{R}^{-1} \tilde{\mathbf{y}}]}$ é uma constante Real e $\int e^{-\frac{1}{2} [(a_2 - \boldsymbol{\Omega} \mathbf{z}'_2 \mathbf{R}^{-1} \tilde{\mathbf{y}})' \boldsymbol{\Omega}^{-1} (a_2 - \boldsymbol{\Omega} \mathbf{z}'_2 \mathbf{R}^{-1} \tilde{\mathbf{y}})]} da_2$ é integral de uma função Gaussiana arbitrária, que admite resultado real, segue de (2.57):

$$p(a_1 | \boldsymbol{\beta}, \boldsymbol{\sigma}_a^2, \boldsymbol{\sigma}_e^2) \propto p(a_1) e^{-\frac{1}{2} [\tilde{\mathbf{y}}' \mathbf{R}^{-1} \tilde{\mathbf{y}} - \tilde{\mathbf{y}}' \mathbf{R}^{-1} \mathbf{z}_2 \boldsymbol{\Omega} \boldsymbol{\Omega}^{-1} \boldsymbol{\Omega} \mathbf{z}'_2 \mathbf{R}^{-1} \tilde{\mathbf{y}}]}$$

Como $p(a_1) = e^{-\frac{a_1 (\boldsymbol{\sigma}_{a_1}^2)^{-1} a_1}{2}}$, temos o desenvolvimento da expressão, ocupando-se das propriedades das variáveis, para, em seguida, estabelecer aspectos de relevância estatística e genética:

$$\begin{aligned}
p(a_1 | \boldsymbol{\beta}, a_2, \boldsymbol{\sigma}_a^2, \boldsymbol{\sigma}_e^2, \mathbf{y}) &\propto e^{-\frac{a_1 (\boldsymbol{\sigma}_{a_1}^2)^{-1} a_1}{2}} e^{-\frac{1}{2} [\tilde{\mathbf{y}}' \mathbf{R}^{-1} \tilde{\mathbf{y}} - \tilde{\mathbf{y}}' \mathbf{R}^{-1} \mathbf{z}_2 \boldsymbol{\Omega} \boldsymbol{\Omega}^{-1} \boldsymbol{\Omega} \mathbf{z}'_2 \mathbf{R}^{-1} \tilde{\mathbf{y}}]} \\
&\propto e^{-\frac{1}{2} [a_1 (\boldsymbol{\sigma}_{a_1}^2)^{-1} a_1 + \tilde{\mathbf{y}}' \mathbf{R}^{-1} \tilde{\mathbf{y}} - \tilde{\mathbf{y}}' \mathbf{R}^{-1} \mathbf{z}_2 \boldsymbol{\Omega} \boldsymbol{\Omega}^{-1} \boldsymbol{\Omega} \mathbf{z}'_2 \mathbf{R}^{-1} \tilde{\mathbf{y}}]} \\
&\propto e^{-\frac{1}{2} [a_1 (\boldsymbol{\sigma}_{a_1}^2)^{-1} a_1 + \tilde{\mathbf{y}}' (\mathbf{R}^{-1} - \mathbf{R}^{-1} \mathbf{z}_2 \boldsymbol{\Omega} \boldsymbol{\Omega}^{-1} \boldsymbol{\Omega} \mathbf{z}'_2 \mathbf{R}^{-1}) \tilde{\mathbf{y}}]} \\
&\propto e^{-\frac{1}{2} [a_1 (\boldsymbol{\sigma}_{a_1}^2)^{-1} a_1 + \tilde{\mathbf{y}}' [\mathbf{R}^{-1} - \mathbf{R}^{-1} \mathbf{z}_2 [\mathbf{z}'_2 \mathbf{R}^{-1} \mathbf{z}_2 - (\boldsymbol{\sigma}_{a_2}^2)^{-1}]^{-1} \mathbf{z}'_2 \mathbf{R}^{-1}] \tilde{\mathbf{y}}]} \\
&\propto e^{-\frac{1}{2} [a_1 (\boldsymbol{\sigma}_{a_1}^2)^{-1} a_1 + \tilde{\mathbf{y}}' [\mathbf{z}_2 \boldsymbol{\sigma}_{a_2}^2 \mathbf{z}'_2 + \mathbf{R}]^{-1} \tilde{\mathbf{y}}]}
\end{aligned}$$

podemos escrever $\mathbf{V}_2 = \mathbf{z}_2 \boldsymbol{\sigma}_{a_2}^2 \mathbf{z}'_2 + \mathbf{R}$ e recordando que $\mathbf{z}_2 \mathbf{z}'_2 = (\mathbf{w}_2 - 2p)(\mathbf{w}_2 - 2p)'$ é o *Kernel* da distância alélica entre os genótipos considerando o segundo SNP. Como o interesse é no primeiro SNP, as operações algébricas seguem recuperando o valor de $\tilde{\mathbf{y}}$ pela mudança de

variável:

$$\begin{aligned}
p(a_1 | \boldsymbol{\beta}, a_2, \sigma_a^2, \sigma_e^2, \mathbf{y}) &\propto e^{-\frac{1}{2} [a_1 (\sigma_{a_1}^2)^{-1} a_1 + (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{z}_1 a_1)' \mathbf{V}_2^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{z}_1 a_1)]} \\
&\propto e^{-\frac{1}{2} [(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \mathbf{V}_2^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) - 2a_1' \mathbf{z}_1' \mathbf{V}_2^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + a_1' \mathbf{z}_1' \mathbf{V}_2^{-1} \mathbf{z}_1 a_1 + a_1' (\sigma_{a_1}^2)^{-1} a_1]} \\
&\propto e^{-\frac{1}{2} [(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \mathbf{V}_2^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) - 2a_1' \mathbf{z}_1' \mathbf{V}_2^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + a_1' [\mathbf{z}_1' \mathbf{V}_2^{-1} \mathbf{z}_1 + (\sigma_{a_1}^2)^{-1}] a_1]}
\end{aligned}$$

denotando $[\mathbf{z}_1' \mathbf{V}_2^{-1} \mathbf{z}_1 + (\sigma_{a_1}^2)^{-1}] = \boldsymbol{\Psi}^{-1}$

$$\begin{aligned}
p(a_1 | \boldsymbol{\beta}, a_2, \sigma_a^2, \sigma_e^2, \mathbf{y}) &\propto e^{-\frac{1}{2} [(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \mathbf{V}_2^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) - 2a_1' \mathbf{z}_1' \mathbf{V}_2^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + a_1' \boldsymbol{\Psi}^{-1} a_1]} \\
&\propto e^{-\frac{1}{2} \left[(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \mathbf{V}_2^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) - 2a_1' \boldsymbol{\Psi}^{-1} \mathbf{z}_1' \mathbf{V}_2^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + a_1' \boldsymbol{\Psi}^{-1} a_1 + \right.} \\
&\quad \left. + (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \mathbf{V}_2^{-1} \mathbf{z}_1 \boldsymbol{\Psi} \boldsymbol{\Psi}^{-1} \mathbf{z}_1' \mathbf{V}_2^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) - (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \mathbf{V}_2^{-1} \mathbf{z}_1 \boldsymbol{\Psi} \boldsymbol{\Psi}^{-1} \mathbf{z}_1' \mathbf{V}_2^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right]} \\
&\propto e^{-\frac{1}{2} \left[(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \mathbf{V}_2^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) - (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \mathbf{V}_2^{-1} \mathbf{z}_1 \boldsymbol{\Psi} \boldsymbol{\Psi}^{-1} \mathbf{z}_1' \mathbf{V}_2^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + \right.} \\
&\quad \left. + (a_1 - \boldsymbol{\Psi} \mathbf{z}_1' \mathbf{V}_2^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}))' \boldsymbol{\Psi}^{-1} (a_1 - \boldsymbol{\Psi} \mathbf{z}_1' \mathbf{V}_2^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})) \right]} \\
&\propto e^{-\frac{1}{2} \left[(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' [\mathbf{V}_2^{-1} - \mathbf{V}_2^{-1} \mathbf{z}_1 \boldsymbol{\Psi} \mathbf{z}_1' \mathbf{V}_2^{-1}] (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + \right.} \\
&\quad \left. + (a_1 - \boldsymbol{\Psi} \mathbf{z}_1' \mathbf{V}_2^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}))' \boldsymbol{\Psi}^{-1} (a_1 - \boldsymbol{\Psi} \mathbf{z}_1' \mathbf{V}_2^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})) \right]} \\
&\propto e^{-\frac{1}{2} \left[(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' [\mathbf{V}_2^{-1} - \mathbf{V}_2^{-1} \mathbf{z}_1 [\mathbf{z}_1' \mathbf{V}_2^{-1} \mathbf{z}_1 + (\sigma_{a_1}^2)^{-1}]^{-1} \mathbf{z}_1' \mathbf{V}_2^{-1}] (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + \right.} \\
&\quad \left. + (a_1 - [\mathbf{z}_1' \mathbf{V}_2^{-1} \mathbf{z}_1 + (\sigma_{a_1}^2)^{-1}]^{-1} \mathbf{z}_1' \mathbf{V}_2^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}))' [\mathbf{z}_1' \mathbf{V}_2^{-1} \mathbf{z}_1 + (\sigma_{a_1}^2)^{-1}] (a_1 - [\mathbf{z}_1' \mathbf{V}_2^{-1} \mathbf{z}_1 + (\sigma_{a_1}^2)^{-1}]^{-1} \mathbf{z}_1' \mathbf{V}_2^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})) \right]} \\
&\propto e^{-\frac{1}{2} \left[(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' [\mathbf{z}_1 \sigma_{a_1}^2 \mathbf{z}_1' + \mathbf{V}_2]^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + \right.} \\
&\quad \left. + (a_1 - [\mathbf{z}_1' \mathbf{V}_2^{-1} \mathbf{z}_1 + (\sigma_{a_1}^2)^{-1}]^{-1} \mathbf{z}_1' \mathbf{V}_2^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}))' [\mathbf{z}_1' \mathbf{V}_2^{-1} \mathbf{z}_1 + (\sigma_{a_1}^2)^{-1}] (a_1 - [\mathbf{z}_1' \mathbf{V}_2^{-1} \mathbf{z}_1 + (\sigma_{a_1}^2)^{-1}]^{-1} \mathbf{z}_1' \mathbf{V}_2^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})) \right]}
\end{aligned}$$

O núcleo é identificado pela expressão da distribuição a posteriori condicional, para o efeito do primeiro SNP, que segue como:

$$p(a_1 | \boldsymbol{\beta}, a_2, \sigma_a^2, \sigma_e^2, \mathbf{y}) \propto e^{-\frac{1}{2} [(a_1 - [\mathbf{z}_1' \mathbf{V}_2^{-1} \mathbf{z}_1 + (\sigma_{a_1}^2)^{-1}]^{-1} \mathbf{z}_1' \mathbf{V}_2^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}))' [\mathbf{z}_1' \mathbf{V}_2^{-1} \mathbf{z}_1 + (\sigma_{a_1}^2)^{-1}] (a_1 - [\mathbf{z}_1' \mathbf{V}_2^{-1} \mathbf{z}_1 + (\sigma_{a_1}^2)^{-1}]^{-1} \mathbf{z}_1' \mathbf{V}_2^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}))]}$$

e, portanto

$$a_1 | \boldsymbol{\beta}, a_2, \sigma_a^2, \sigma_e^2, \mathbf{y} \sim N \left[(\mathbf{z}_1' \mathbf{V}_2^{-1} \mathbf{z}_1 + (\sigma_{a_1}^2)^{-1})^{-1} \mathbf{z}_1' \mathbf{V}_2^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}), [\mathbf{z}_1' \mathbf{V}_2^{-1} \mathbf{z}_1 + (\sigma_{a_1}^2)^{-1}]^{-1} \right] \quad (2.58)$$

No segundo estágio da hierarquia, as funções distribuição a priori, para os hiperparâmetros (componentes da variância, auxiliam na especificação das prioris e distingue dos parâmetros de interesse) das funções assumidas, no primeiro estágio foram especificadas por meio da distribuição qui-quadrado invertida escalonada, $\sigma_{a_i}^2 | v_i, s_i^2 \sim \chi^{-2}(v_i, s_i^2)$, com parâmetro escala s_i^2 e função densidade:

$$p(\sigma_{a_i}^2 | v_i, s_i^2) \propto (\sigma_{a_i}^2)^{-\left(\frac{v_i+2}{2}\right)} \exp\left(\frac{-v_i s_i^2}{2\sigma_{a_i}^2}\right) \quad i = 1, \dots, k \quad (2.59)$$

v_i são os graus de liberdade e o valor inicial da variância $\sigma_{a_i}^2$.

Definidas estas distribuições, pode-se apresentar a distribuição conjunta a posteriori do parâmetro σ_k^2 , dado efeito de k marcadores do modelo e obtida pelo produto de todas as densidades específicas:

$$\begin{aligned} p(\sigma_k^2 | a_1, \dots, a_k) &\propto p(a_1, \dots, a_k | \sigma_k^2) p(\sigma_k^2) \\ &\propto p(a_1) \dots p(a_k) p(\sigma_k^2) \\ &\propto \prod_{i=1}^k p(a_i) \cdot p(\sigma_k^2) \\ &\propto 2\pi (\sigma_k^2)^{-\frac{1}{2}} e^{-\frac{1}{2} a_1' (\sigma_k^2)^{-1} a_1} \dots 2\pi (\sigma_k^2)^{-\frac{1}{2}} e^{-\frac{1}{2} a_k' (\sigma_k^2)^{-1} a_k} \cdot (\sigma_k^2)^{-\left(\frac{v_k+2}{2}\right)} e^{\left(\frac{-v_k s_k^2}{2\sigma_k^2}\right)} \\ &\propto 2\pi^k (\sigma_k^2)^{-\frac{k}{2}} e^{-\frac{1}{2\sigma_k^2} (a_1' a_1 + \dots + a_k' a_k)} \cdot (\sigma_k^2)^{-\left(\frac{v_k+2}{2}\right)} e^{\left(\frac{-v_k s_k^2}{2\sigma_k^2}\right)} \\ &\propto (\sigma_k^2)^{-\left(\frac{k}{2} + \frac{v_k}{2} + 1\right)} e^{-\frac{1}{2\sigma_k^2} (\mathbf{a}'\mathbf{a} + v_k s_k^2)} \end{aligned}$$

Portanto $\sigma_k^2 | \mathbf{a} \sim \chi^{-2}(v_k + k, \mathbf{a}'\mathbf{a} + v_k s_k^2)$ é a distribuição qui-quadrado invertida escalonada com $\mathbf{a}'\mathbf{a} + v_k s_k^2$ graus de liberdade e $v_k + k$ parâmetro escala.

2.5.1 Recuperação da Informação de Parentesco e Inversão da matriz V_2

Para populações em Equilíbrio de Hardy-Weinberg (HW), utilizando os valores genotípicos codificados, a média e a variância genotípica (σ_g^2) são consideradas:

$$E(w_{ij}) = (p - q)a + 2pqd$$

aqui, a e d são o efeito aditivo e desvio de dominância, respectivamente, p e q são as frequências dos alelos.

$$\text{Var}(w_{ij}) = \sigma_g^2 = \underbrace{2pq\alpha^2}_{\sigma_A^2} + \underbrace{(2pqd)^2}_{\sigma_D^2}$$

nos quais σ_A^2 é a variância genética aditiva, σ_D^2 é a variância de dominância e α é efeito da substituição alélica, dado por:

$$\alpha = [a - (p - q)d]$$

Para tanto, ao destacarmos a variância genética aditiva para o locus j , temos que

$$\sigma_{A_j}^2 = 2p_jq_j\alpha_j^2 \Rightarrow E(\sigma_{A_j}^2) = 2p_jq_jE(\alpha_j^2)$$

assumindo que $E(\alpha_j^2) \approx \sigma_{a_j}^2$ para uma distribuição qui-quadrado invertida escalonada então

$$\sigma_{A_j}^2 \approx 2p_jq_j\sigma_{a_j}^2$$

Assim sendo:

$$\frac{\sigma_{A_j}^2}{2p_jq_j} \approx \sigma_{a_j}^2$$

Agora, voltando a observar a distribuição (2.58), no caso em que $j = 2$, destacamos o *Kernel* da variância obtido:

$$\mathbf{V}_2 = \mathbf{z}_2\sigma_{a_2}^2\mathbf{z}_2' + \mathbf{R} \quad (2.60)$$

podemos substituir $\sigma_{a_2}^2 = \frac{\sigma_{A_2}^2}{2p_2q_2}$

$$\begin{aligned} \mathbf{V}_2 &= \mathbf{z}_2\frac{\sigma_{A_2}^2}{2p_2q_2}\mathbf{z}_2' + \mathbf{R} \\ \mathbf{V}_2 &= \frac{(\mathbf{w}_2 - 2p_2)(\mathbf{w}_2 - 2p_2)'}{2p_2q_2}\sigma_{A_2}^2 + \mathbf{R} \end{aligned} \quad (2.61)$$

em que temos $\frac{(\mathbf{w}_2 - 2p_2)(\mathbf{w}_2 - 2p_2)'}{2p_2q_2}\sigma_{A_2}^2$ exatamente o valor de parentesco genômico unificado de Powell 2010 (POWELL; GODDARD, 2010). Se considerarmos todos os loci, a matriz de Powell é ponderada pela variância desigual, ao longo de cada locus, mas se é assumida variância

igual, como em modelos mistos, então, a matriz de covariância torna-se:

$$\frac{(\mathbf{z}_{k-1})(\mathbf{z}_{k-1})'}{2 \sum_{j=1}^{k-1} p_j q_j} \sigma_{A_{k-1}}^2 + \mathbf{R}$$

Usando essa covariância genômica entre os genótipos, levamos em consideração a estrutura da população. Nota-se que, na abordagem de modelo misto apresentada por Lippert et al. (2011), Xu (2013) e Yu e Buckler (2006), temos $\mathbf{z}_{k-1}\mathbf{z}'_{k-1} = \mathbf{z}\mathbf{z}'$ e $\sigma_{A_{k-1}}^2 = \sigma_A^2$. Portanto o marcador ajustado compartilha informações também do efeito poligênico. Em nossa abordagem, este não é o caso.

Uma questão agora passa a ser a necessidade de inverter a matriz \mathbf{V}_2^{-1} , enquanto, em mapeamento associativo é clássico o mapeamento componentes a fim de evitar várias inversões em múltiplas varreduras, neste contexto a transformação de dados por componentes principais não são necessários. Em vez disso, usa-se a transformação: Fazendo $\mathbf{V}_2 = \mathbf{V} - \mathbf{z}_1\mathbf{z}'_1\sigma_{a_1}^2$ e multiplicando por $\mathbf{z}'_1\mathbf{z}_1$ e $\mathbf{z}'_1\tilde{\mathbf{y}}$ ambos os lados da matriz inversa, tem-se, respectivamente:

$$\begin{aligned} \mathbf{z}'_1\mathbf{V}_2^{-1}\mathbf{z}_1 &= \mathbf{z}'_1[\mathbf{V} - \mathbf{z}_1\mathbf{z}'_1\sigma_{a_1}^2]^{-1}\mathbf{z}_1 = \mathbf{z}'_1\mathbf{V}^{-1}\mathbf{z}_1 - \mathbf{z}'_1\mathbf{V}^{-1}\mathbf{z}_1[\mathbf{z}'_1\mathbf{V}^{-1}\mathbf{z}_1 - (\sigma_{a_1}^2)^{-1}]^{-1}\mathbf{z}'_1\mathbf{V}^{-1}\mathbf{z}_1 \\ \mathbf{z}'_1\mathbf{V}_2^{-1}\tilde{\mathbf{y}} &= \mathbf{z}'_1[\mathbf{V} - \mathbf{z}_1\mathbf{z}'_1\sigma_{a_1}^2]^{-1}\tilde{\mathbf{y}} = \mathbf{z}'_1\mathbf{V}^{-1}\tilde{\mathbf{y}} - \mathbf{z}'_1\mathbf{V}^{-1}\mathbf{z}_1[\mathbf{z}'_1\mathbf{V}^{-1}\mathbf{z}_1 - (\sigma_{a_1}^2)^{-1}]^{-1}\mathbf{z}'_1\mathbf{V}^{-1}\tilde{\mathbf{y}} \end{aligned}$$

Assumindo que $\phi = \mathbf{z}'_1\mathbf{V}_2^{-1}\mathbf{z}_1$ e $\omega = \mathbf{z}'_1\mathbf{V}_2^{-1}\tilde{\mathbf{y}}$

$$\begin{aligned} \mathbf{z}'_1\mathbf{V}_2^{-1}\mathbf{z}_1 &= \phi - \phi[\phi - (\sigma_{a_1}^2)^{-1}]^{-1}\phi \\ \mathbf{z}'_1\mathbf{V}_2^{-1}\tilde{\mathbf{y}} &= \omega - \phi[\phi - (\sigma_{a_1}^2)^{-1}]^{-1}\omega \end{aligned}$$

Agora, substituindo na equação da distribuição condicional completa posteriori (2.58), segue

$$a_1|\boldsymbol{\beta}, a_2, \sigma_a^2, \sigma_e^2, \mathbf{y} \sim N \left[(\phi - \phi\xi^{-1} + (\sigma_{a_1}^2)^{-1})^{-1} (\omega - \phi\xi^{-1}\omega), (\phi - \phi\xi^{-1} + (\sigma_{a_1}^2)^{-1})^{-1} \right] \quad (2.62)$$

em que $\xi^{-1} = [\phi - (\sigma_{a_1}^2)^{-1}]^{-1}$

3 MATERIAL E MÉTODOS

O estudo está subdividido em quatro etapas principais. Primeiro, os dados genotípicos de três populações de indivíduos foram simulados. Em segundo lugar, os efeitos fenotípicos foram simulados a partir de todo o conjunto de SNPs. O terceiro passo fundamentou-se em avaliar o desempenho dos métodos propostos e compará-lo aos métodos RR-BLUP, Bayes A, Bayes B, BSSV, ao método GWAS proposto por Yu et al. (2006), além dos métodos Multiloci FASTmrMLM e FASTmrEMMA. A quarta etapa aplicou-se à análise a dados de 230 linhagens de milho genotipadas, com 23.153 marcadores, elucidando as propriedades estatísticas e adaptações de cada um, por meio de comparações entre os métodos e aplicação do teste de Wald.

3.1 Material

3.1.1 Simulação do Genótipo

Na condução da simulação para obter os dados genotípicos de três populações dialélicas, utilizou-se o programa GENES (CRUZ, 2013). O programa GENES efetua análises, procedimentos de estatística experimental, multivariada, melhoramento genético, além de simular dados genéticos, utilizando diferentes tipos e tamanhos de populações, o que permite valores para a análise de mapeamento, dentre outros estudos genéticos.

Os parâmetros e valores para simular as três populações dialélicas:

- número de populações: 3
- número de indivíduos/população: 200
- número de loci: 6000
- Tipo de marcador: Co-dominante

Figura 3.1 – Arquivo de dados simulados pelo Programa GENES

| Programa | GENES | Simulação | de | Populações e | Marcadores | Dominantes/ |
|----------|------------|----------------------|----|--------------|------------|---------------|
| Arquivo | de | saída | de | Simular1 | | Co-dominantes |
| Número | de | populações | | 3 | | |
| Número | de | indivíduos/população | | 200 | | |
| Variação | de | locos | 0 | | | |
| Número | de | | | 6000 | | |
| Data | 17/11/2016 | | | | | |

| Pop | Loco | N.alelo | Frequência | | | | | | | | | |
|-----|------|---------|------------|--|--|--|--|--|--|--|--|--|
| 1 | 1 | 2 | 35 | | | | | | | | | |
| 1 | 2 | 2 | 86 | | | | | | | | | |
| 1 | 3 | 2 | 14 | | | | | | | | | |
| ... | ... | ... | ... | | | | | | | | | |
| 3 | 6000 | 2 | 44 | | | | | | | | | |

| | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 1 | 22 | 22 | 12 | 22 | 22 | 11 | 12 | 22 | 22 | ... | 11 |
| 1 | 22 | 22 | 12 | 22 | 22 | 11 | 12 | 12 | 22 | ... | 11 |
| 1 | 11 | 22 | 12 | 22 | 22 | 11 | 22 | 22 | 22 | ... | 12 |
| 1 | 22 | 22 | 12 | 22 | 22 | 11 | 12 | 12 | 22 | ... | 11 |
| 1 | 22 | 22 | 11 | 12 | 22 | 11 | 22 | 12 | 22 | ... | 11 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 2 | 22 | 12 | 11 | 22 | 22 | 22 | 22 | 22 | 22 | ... | 22 |
| 2 | 22 | 12 | 11 | 12 | 22 | 12 | 22 | 12 | 22 | ... | 11 |
| 2 | 22 | 22 | 12 | 22 | 22 | 11 | 12 | 22 | 22 | ... | 11 |
| 2 | 22 | 22 | 12 | 22 | 22 | 12 | 22 | 22 | 22 | ... | 12 |
| 2 | 22 | 22 | 11 | 22 | 22 | 11 | 22 | 12 | 22 | ... | 11 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 3 | 22 | 22 | 11 | 22 | 22 | 12 | 22 | 22 | 12 | ... | 22 |
| 3 | 22 | 22 | 11 | 22 | 12 | 11 | 12 | 11 | 22 | ... | 11 |
| 3 | 22 | 22 | 11 | 12 | 22 | 11 | 22 | 12 | 22 | ... | 11 |
| 3 | 12 | 22 | 12 | 22 | 12 | 11 | 22 | 22 | 22 | ... | 11 |
| 3 | 22 | 22 | 11 | 12 | 22 | 11 | 22 | 12 | 22 | ... | 11 |

Fonte: Da autora (2019)

Ao realizar a simulação dos marcadores e estratificação da população, em três grupos genotípicos, permitiu-se a codificação dos SNPs em variáveis categorizadas 0, 1 e 2 de forma arbitrária, indicando os genótipos MM, Mm e mm (ou 22, 12 e 11 conforme apresentado na Figura 3.1) e consideraram-se os valores genéticos a , d e $-a$, respectivamente. Assim a matriz segue

$$\mathbf{W} = [w_{ij}]_{n \times k}, \text{ em que } w_{ij} = \begin{cases} 0, & \text{com probabilidade } p^2; \\ 1, & \text{com probabilidade } 2pq, \quad i = 1, \dots, n \\ 2, & \text{com probabilidade } q^2, \quad j = 1, \dots, k \end{cases}$$

A matriz $\mathbf{W} = [w_{ij}]_{n \times k}$ obtida com os dados simulados constituiu-se de $n=600$ indivíduos (200 indivíduos por população), com $k=6000$ marcadores do tipo SNPs co-dominantes,

distribuídos sobre 6 cromossomos (1000 marcadores SNPs por cromossomo), segue

$$\mathbf{W} = \begin{bmatrix} \mathbf{W}_1 \\ \mathbf{W}_2 \\ \mathbf{W}_3 \end{bmatrix} = \begin{bmatrix} w_{11} & \dots & w_{1k} \\ \vdots & \ddots & \vdots \\ w_{2001} & \dots & w_{200k} \\ \hline w_{2011} & \dots & w_{201k} \\ \vdots & \ddots & \vdots \\ w_{4001} & \dots & w_{400k} \\ \hline w_{4011} & \dots & w_{401k} \\ \vdots & \ddots & \vdots \\ w_{6001} & \dots & w_{600k} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & \dots & 0 & 2 & \dots & 2 & \dots \\ 0 & 2 & 1 & \dots & 1 & 0 & \dots & 1 & \dots \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots & \dots & \vdots & \dots \\ 1 & 1 & 0 & \dots & 0 & 1 & \dots & 0 & \dots \\ \hline 2 & 0 & 2 & \dots & 0 & 0 & \dots & 2 & \dots \\ 0 & 0 & 1 & \dots & 1 & 1 & \dots & 0 & \dots \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots & \dots & \vdots & \dots \\ 1 & 1 & 0 & \dots & 0 & 2 & \dots & 2 & \dots \\ \hline 2 & 2 & 2 & \dots & 0 & 0 & \dots & 2 & \dots \\ 0 & 1 & 0 & \dots & 0 & 0 & \dots & 1 & \dots \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots & \dots & \vdots & \dots \\ 1 & 2 & 0 & \dots & 0 & 1 & \dots & 0 & \dots \end{bmatrix}_{600 \times 6000}$$

no qual \mathbf{W} pode ser vista de forma particionada em matrizes \mathbf{W}_1 , \mathbf{W}_2 e \mathbf{W}_3 . Cada matriz faz corresponder aos genótipos da População 1, 2 e 3, respectivamente e cada população com 200 indivíduos genotipados na ordem de 6000 SNPs. Estes genótipos formam a base para a simulação dos fenótipos.

3.1.2 Simulação dos Fenótipos

Com o uso do *software* R-3.4.3 (64 bit) (R DEVELOPMENT CORE TEAM, 2016), foram inicialmente simulados valores fenotípicos ($\mathbf{y}_{n \times 1}$) de um modelo linear genético

$$\mathbf{y} = (\mathbf{W} - 2p\mathbf{J})\alpha \quad (3.1)$$

em que

\mathbf{y} vetor ($n \times 1$) de fenótipos, $n = 600$ é o número de observações;

\mathbf{W} matriz ($n \times k$) genotípica de SNPs, $k = 6000$ é o número de SNPs simulados;

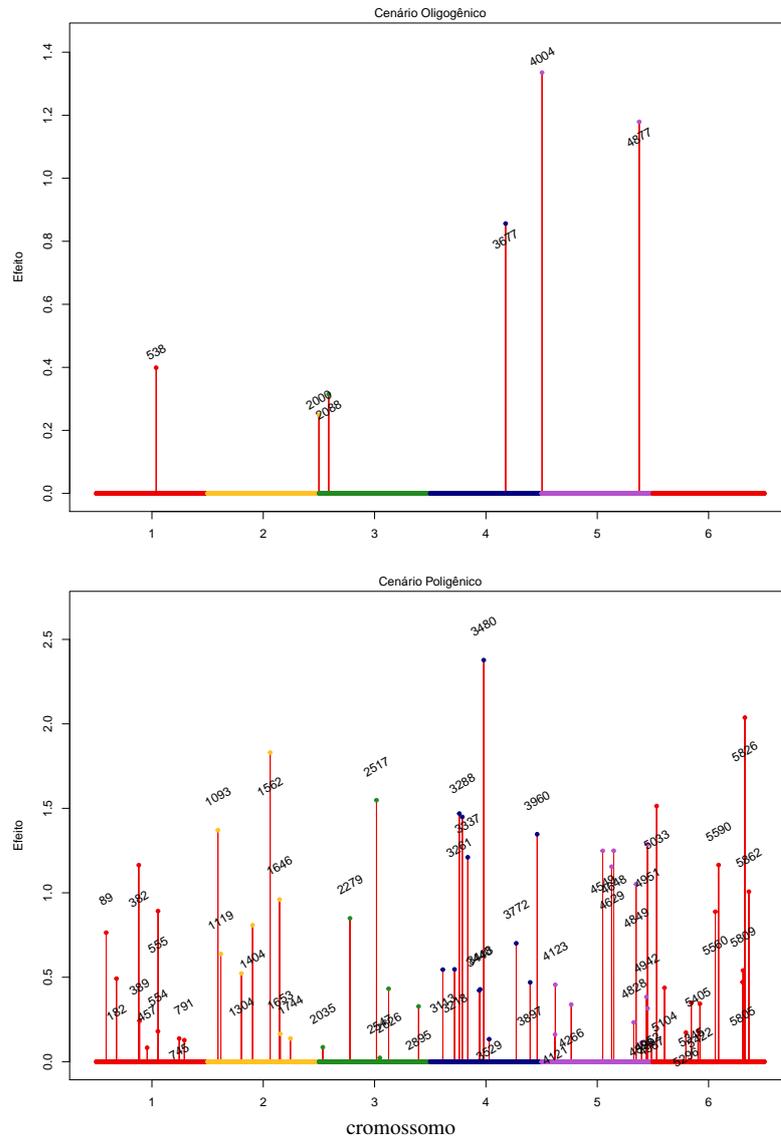
$2p$ média da variável w_{ij} associada à matriz de incidência \mathbf{W} ;

J matriz ($n \times k$) cujos elementos são todos iguais e iguais a um;

α efeito da substituição alélica, simulado de uma distribuição $N(0, 1)$.

As herdabilidades $h^2 \in \{0, 2; 0, 5; 0, 8\}$ foram consideradas para compor dois cenários: oligogênico e poligênico. Os aspectos de cada cenário relacionaram o valor dos efeitos genéticos de cada locus e criaram arquiteturas genéticas diferentes. Geraram-se dois vetores. Um representou uma fração de 0,1% dos 6000 marcadores com efeito, para o cenário oligogênico (apresentam herança controlada por poucos genes) e 1% dos 6000 marcadores com efeito simulado representou o cenário poligênico (refere-se às características fenotípicas determinadas por muitos genes). Os QTLs (*Quantitative trait locus*) foram obtidos aleatoriamente, com efeitos simulados de uma distribuição normal padrão. A Figura 3.2 exibe painel com os respectivos cenários.

Figura 3.2 – Painéis demonstrando as marcas dos efeitos simulados em cada cenário Oligogênico (6 genes) e Poligênico (60 genes)



Fonte: Da autora (2019)

Adotou-se um fenótipo específico com base nos efeitos que compõem as fontes de variação de cada coeficiente de herdabilidade em cada cenário. Os modelos foram: oligogênico herdabilidade 0,2 (yo2), oligogênico herdabilidade 0,5 (yo5), oligogênico herdabilidade 0,8 (yo8), poligênico herdabilidade 0,2 (yp2), poligênico herdabilidade 0,5 (yp5) e poligênico herdabilidade 0,8 (yp8).

As simulações para os seis fenótipos, avançaram com uso do *software* R-3.4.3 (64 bit) (R DEVELOPMENT CORE TEAM, 2016) e o *script* para a obtenção dos fenótipos segue em anexo.

3.1.3 Dados de resistência à podridão de espiga causada por *Fusarium verticillioides*

Nos dados do experimento de milho, depois de avaliadas 230 linhagens em análise conjunta (locais Lavras e Uberlândia) para a proporção de espigas com sintomas de podridão, o material genético utilizado consistiu de 23153 marcadores do tipo SNPs derivados da metodologia DART-seq (*Diversity Arrays Technology*), num total de 10 cromossomos identificados e, a partir da posição 11, remete como posição de cromossomos não identificados.

Os experimentos foram conduzidos ,utilizando o delineamento experimental de blocos incompletos com três repetições por local e compostos por uma linha de 3 metros com espaçamento de 0,7 metros.

3.2 Métodos

3.2.1 Modelo Linear Misto para Análise Bayesiana

Considerou-se o modelo linear misto clássico a fim de definir os métodos modificados

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{a} + \mathbf{e} \quad (3.2)$$

nos quais \mathbf{y} é o vetor $n \times 1$ de observações (fenótipos), \mathbf{X} é uma matriz $n \times p$ de incidência relacionada ao vetor $\boldsymbol{\beta}$ de tamanho $p \times 1$ e efeito fixo para \mathbf{y} , \mathbf{Z} é uma matriz $n \times k$ de covariáveis genóticas (codificadas como 0, 1 ou 2) para k marcadores do tipo SNPs. \mathbf{a} é vetor $k \times 1$ de efeito aleatórios dos k SNPs (comumente conhecidos como efeito dos marcadores) e \mathbf{e} $n \times 1$ é o vetor de efeito residual. Como o vetor é dado por $\mathbf{a}_{k \times 1} = (a_1, a_2, \dots, a_k)'$ e a matriz de incidência do efeito marcadores pode ser escrita como $\mathbf{Z}_{n \times k} = \begin{bmatrix} \mathbf{z}_1 & \mathbf{z}_2 & \dots & \mathbf{z}_k \end{bmatrix}$, segue (3.2) reescrita como:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{z}_1 a_1 + \mathbf{z}_2 a_2 + \dots + \mathbf{z}_k a_k + \mathbf{e} \quad (3.3)$$

Para a análise Bayesiana do modelo linear misto 3.3, os elementos considerados são as informações a priori sobre os parâmetros utilizados em associação com os dados amostrais (pelo teorema de Bayes, representa a função de verossimilhança), o que promove a inferência baseada na distribuição a posteriori. O vetor de efeito residual pressupõe de uma distribuição

normal com média zero e variância $\mathbf{R} = \mathbf{I}\sigma_e^2$, em que \mathbf{I} é a matriz identidade.

$$\mathbf{e} \sim N(\mathbf{0}, \mathbf{I}\sigma_e^2)$$

O vetor de observações \mathbf{y} segue a distribuição condicional normal multivariada, tal qual

$$\mathbf{y}|\boldsymbol{\beta}, \mathbf{a}, \sigma_a^2, \sigma_e^2 \sim N(\mathbf{X}\boldsymbol{\beta} + \mathbf{z}_1 a_1 + \dots + \mathbf{z}_k a_k, \mathbf{I}\sigma_e^2) \quad (3.4)$$

Assim, a função de verossimilhança é dada por:

$$\begin{aligned} p(\mathbf{y}|\boldsymbol{\beta}, \mathbf{a}, \sigma_a^2, \sigma_e^2) &= \left(\frac{1}{2\pi\sigma_e^2} \right)^{\frac{n}{2}} e^{-\frac{1}{2} \frac{(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}_1 a_1 - \dots - \mathbf{Z}_k a_k)' (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}_1 a_1 - \dots - \mathbf{Z}_k a_k)}{\sigma_e^2}} \\ &\propto e^{-\frac{1}{2} \frac{(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}_1 a_1 - \dots - \mathbf{Z}_k a_k)' (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}_1 a_1 - \dots - \mathbf{Z}_k a_k)}{\sigma_e^2}} \end{aligned} \quad (3.5)$$

3.2.2 Distribuições a Priori

Ao prosseguir com a análise Bayesiana, as distribuições a priori foram especificadas para $\boldsymbol{\beta}$ e \mathbf{a} . Em todos os métodos, considerou-se $\boldsymbol{\beta}$ seguindo uma distribuição a priori (*flat*). A suposição de que não haja informação prévia a respeito dos efeitos fixos é comum. Sendo assim, a priori de $\boldsymbol{\beta}$, nestes casos, foi uma candidata natural de distribuição a priori imprópria (SORENSEN; GIANOLA, 2007).

$$p(\boldsymbol{\beta}) \propto \text{constante} \quad (3.6)$$

Também admitiu-se a distribuição a priori para cada efeito aleatório dos marcadores:

$$a_i \sim N(0, \sigma_{a_i}^2) \quad i = 1, \dots, k$$

Desta forma, a distribuição, para os efeitos de marcadores condicionais, as variâncias específicas de efeito de marca são independentes e normalmente distribuídas com função da densidade

$$p(\mathbf{a}|\boldsymbol{\sigma}_a^2) = \prod_{i=1}^k (2\pi\sigma_{a_i}^2)^{-\frac{1}{2}} \exp\left[-\frac{a_i^2}{2\sigma_{a_i}^2}\right] \quad i = 1, \dots, k \quad (3.7)$$

na qual $\boldsymbol{\sigma}_a^2 = (\sigma_{a_1}^2, \sigma_{a_2}^2, \dots, \sigma_{a_k}^2)$ é um vetor $k \times 1$, com variância para cada marcador específico.

No segundo estágio da hierarquia, as funções a priori, para cada $\sigma_{a_i}^2$, $i = 1, \dots, k$ foram idênticas e independentes, seguiram a distribuição qui-quadrado invertida escalonada, com parâmetro escala s_i^2 e v_i graus de liberdade.

$$\sigma_{a_i}^2 | v_i, s_i^2 \sim \chi^{-2}(v_i, s_i^2) \quad \text{ou seja} \quad p(\sigma_{a_i}^2 | v_i, s_i^2) \propto (\sigma_{a_i}^2)^{-\left(\frac{v_i+2}{2}\right)} \exp\left(\frac{-v_i s_i^2}{2\sigma_{a_i}^2}\right) \quad (3.8)$$

tal que

$$p(\boldsymbol{\sigma}_a^2 | v_i, s_i^2) = \prod_{i=1}^k \frac{(v_i s_i^2 / 2)^{\frac{v_i}{2}}}{\Gamma(v_i/2)} (\sigma_{a_i}^2)^{-\left(\frac{v_i+2}{2}\right)} \exp\left(\frac{-v_i s_i^2}{2\sigma_{a_i}^2}\right) \quad (3.9)$$

e a função densidade a priori, para a variância residual σ_e^2 , segue a distribuição qui-quadrado invertida escalonada, com parâmetro escala s_e^2 e v_e graus de liberdade

$$\sigma_e^2 | v_e, s_e^2 \sim \chi^{-2}(v_e, s_e^2) \quad \text{ou seja} \quad p(\sigma_e^2 | v_e, s_e^2) \propto (\sigma_e^2)^{-\left(\frac{v_e+2}{2}\right)} \exp\left(\frac{-v_e s_e^2}{2\sigma_e^2}\right) \quad (3.10)$$

Os valores a priori dos hiperparâmetros foram considerados, conforme descritos por Garrick e Fernando (2013) e assumiram parâmetro escala $v_0 = 1$ e $s_0^2 = \frac{\sigma_y^2 0,005}{k}$ graus de liberdade.

A distribuição conjunta a posteriori do parâmetro σ_k^2 , dado efeito de k -ésimo marcador do modelo, obtida pelo produto de todas as densidades específicas segue

$$p(\sigma_k^2 | a_1, \dots, a_k) \propto (\sigma_k^2)^{-\left(\frac{k}{2} + \frac{v_k}{2} + 1\right)} e^{-\frac{1}{2\sigma_k^2}(\mathbf{a}'\mathbf{a} + v_k s_k^2)} \quad (3.11)$$

ou seja, $\sigma_k^2 | \mathbf{a} \sim \chi^{-2}(v_k + k, \mathbf{a}'\mathbf{a} + v_k s_k^2)$ segue a distribuição qui-quadrado invertida escalonada com $v_k + k$ graus de liberdade e parâmetro escala $\mathbf{a}'\mathbf{a} + v_k s_k^2$.

3.2.3 Distribuição a Posteriori Conjunta

Nas análises bayesianas, as inferências dos parâmetros foram baseadas em distribuições a posteriori conjunta, sendo, na maioria das vezes, de expressão complexa, o que dificulta a obtenção analítica das distribuições marginais de interesse na análise.

A distribuição a posteriori conjunta de todos os parâmetros é resultado do produto de todas as distribuições a priori e a verossimilhança (consideradas anteriormente). Portanto a distribuição a posteriori conjunta foi dada pela seguinte expressão, supondo-se independência

entre os parâmetros

$$\begin{aligned}
p(\boldsymbol{\beta}, \mathbf{a}, \boldsymbol{\sigma}_a^2, \sigma_e^2 | \mathbf{y}) &\propto p(\boldsymbol{\beta})p(\mathbf{a})p(\boldsymbol{\sigma}_a^2)p(\sigma_e^2)p(\mathbf{y} | \boldsymbol{\beta}, \mathbf{a}, \boldsymbol{\sigma}_a^2, \sigma_e^2) \\
&\propto p(\boldsymbol{\beta})p(\mathbf{a})p(\boldsymbol{\sigma}_a^2)p(\sigma_e^2)e^{-\frac{1}{2} \frac{(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}_1\mathbf{a}_1 - \dots - \mathbf{Z}_k\mathbf{a}_k)'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}_1\mathbf{a}_1 - \dots - \mathbf{Z}_k\mathbf{a}_k)}{\sigma_e^2}} \quad (3.12)
\end{aligned}$$

As inferências sobre cada parâmetro, geralmente, é feita utilizando as distribuições marginais. No entanto as expressões de forma fechada, obtidas a partir da expressão 3.12, não foram possíveis para tais inferências. Assim, pode-se inferir sobre um único parâmetro de interesse, sendo os demais conhecidos como parâmetros perturbadores. As amostras foram retiradas das distribuições a posteriori condicionais, para cada efeito de marcador, com propósito de implementar o amostrador de Gibbs, o qual permite ter aproximações das distribuições marginais.

3.2.4 Marginalização dos Parâmetros Perturbadores

Considerada a identificação da função distribuição a posteriori condicional completa, para efeito de cada marcador do tipo SNP, a função $p(a_i | \boldsymbol{\beta}, \mathbf{a}_{-i}, \boldsymbol{\sigma}_a^2, \sigma_e^2, \mathbf{y})$ foi escrita como

$$\begin{aligned}
p(a_i | \boldsymbol{\beta}, \mathbf{a}_{-i}, \boldsymbol{\sigma}_a^2, \sigma_e^2, \mathbf{y}) &= \frac{p(\boldsymbol{\beta}, a_i, \mathbf{a}_{-i}, \boldsymbol{\sigma}_a^2, \sigma_e^2 | \mathbf{y})}{p(\boldsymbol{\beta}, \mathbf{a}_{-i}, \boldsymbol{\sigma}_a^2, \sigma_e^2 | \mathbf{y})} \\
&\propto p(\boldsymbol{\beta}, a_i, \mathbf{a}_{-i}, \boldsymbol{\sigma}_a^2, \sigma_e^2 | \mathbf{y}) \\
&\propto L(a_i | \boldsymbol{\beta}, \mathbf{a}_{-i}, \boldsymbol{\sigma}_a^2, \sigma_e^2, \mathbf{y})p(a_i)p(\mathbf{a}_{-i}) \quad (3.13)
\end{aligned}$$

a relevância na equação 3.13 corresponde à função de verossimilhança $L(a_i | \boldsymbol{\beta}, \mathbf{a}_{-i}, \boldsymbol{\sigma}_a^2, \sigma_e^2, \mathbf{y})$, que trata \mathbf{a}_{-i} como constante. As funções $p(a_i)$ e $p(\mathbf{a}_{-i})$ são as funções densidades da distribuição a priori de a_i e \mathbf{a}_{-i} , respectivamente. A partir deste ponto, operou-se com a integração da distribuição a posteriori condicional, varrendo o domínio de $(k - 1)$ parâmetros perturbadores, declarado como (\mathbf{a}_{-i}) . O interesse sendo apenas o i -ésimo SNP, foi marginalizado em relação

aos outros efeitos ($k - 1$) SNPs, concedido pela expressão

$$\begin{aligned}
& p(a_i | \boldsymbol{\beta}, \mathbf{a}_{-i}, \boldsymbol{\sigma}_a^2, \boldsymbol{\sigma}_e^2, \mathbf{y}) \propto \\
& \propto (\boldsymbol{\sigma}_{a_i}^2)^{-\frac{1}{2}} e^{-\frac{a_i(\boldsymbol{\sigma}_{a_i}^2)^{-1}a_i}{2}} \iiint \dots \int e^{-\frac{\left[\tilde{\mathbf{y}} - \sum_{j=1}^{k-1} \mathbf{z}_j a_j - \mathbf{z}_i a_i \right] \mathbf{R}^{-1} \left[\tilde{\mathbf{y}} - \sum_{j=1}^{k-1} \mathbf{z}_j a_j - \mathbf{z}_i a_i \right]}{2}} \prod_{j=1}^{k-1} (\boldsymbol{\sigma}_{a_j}^2)^{-\frac{1}{2}} e^{-\frac{a_j(\boldsymbol{\sigma}_{a_j}^2)^{-1}a_j}{2}} da_j \\
& \propto (\boldsymbol{\sigma}_{a_i}^2)^{-\frac{1}{2}} e^{-\frac{a_i(\boldsymbol{\sigma}_{a_i}^2)^{-1}a_i}{2}} e^{-\frac{(\tilde{\mathbf{y}} - \mathbf{z}_i a_i)' \left[\sum_{j=1}^{k-1} \mathbf{z}_j \mathbf{z}_j' \boldsymbol{\sigma}_{a_j}^2 + \mathbf{R} \right]^{-1} (\tilde{\mathbf{y}} - \mathbf{z}_i a_i)}{2}} \\
& \propto (\boldsymbol{\sigma}_{a_i}^2)^{-\frac{1}{2}} e^{-\frac{a_i(\boldsymbol{\sigma}_{a_i}^2)^{-1}a_i}{2}} e^{-\frac{(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{z}_i a_i)' [\mathbf{z}_{k-1} \boldsymbol{\Psi} \mathbf{z}_{k-1}' + \mathbf{R}]^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{z}_i a_i)}{2}} \tag{3.14}
\end{aligned}$$

no que diz respeito à \mathbf{Z}_{k-1} , de dimensão $n \times (k - 1)$, é a matriz de SNPs tal que a informação do i -ésimo marcador é ausente. A matriz diagonal $\boldsymbol{\Psi} = [\mathbf{z}_i' \mathbf{V}_{k-1}^{-1} \mathbf{z}_i + (\boldsymbol{\sigma}_{a_i}^2)^{-1}]^{-1}$ assume variância específica para cada marcador. Portanto segue

$$a_i | \boldsymbol{\beta}, \mathbf{a}_{-i}, \boldsymbol{\sigma}_{a_{k-1}}^2, \boldsymbol{\sigma}_e^2, \mathbf{y} \sim N \left[(\mathbf{z}_i' \mathbf{V}_{k-1}^{-1} \mathbf{z}_i + (\boldsymbol{\sigma}_{a_i}^2)^{-1})^{-1} \mathbf{z}_i' \mathbf{V}_{k-1}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}), [\mathbf{z}_i' \mathbf{V}_{k-1}^{-1} \mathbf{z}_i + (\boldsymbol{\sigma}_{a_i}^2)^{-1}]^{-1} \right] \tag{3.15}$$

O desenvolvimento analítico, para a obtenção das distribuições encontra-se no Apêndice A.

Para o vetor $\boldsymbol{\beta}$ de efeitos fixos não genéticos, a distribuição a posteriori condicional completa é normal com média e variância que seguem

$$\boldsymbol{\beta} | \mathbf{a}, \boldsymbol{\sigma}_a^2, \boldsymbol{\sigma}_e^2, \mathbf{y} \sim N \left[(\mathbf{X}' \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}' \mathbf{V}^{-1} (\mathbf{y}), [\mathbf{X}' \mathbf{V}^{-1} \mathbf{X}]^{-1} \right] \tag{3.16}$$

decorre que $\mathbf{V} = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R}$. A priori de $\boldsymbol{\beta}$ admitida como *flat* e, portanto a função densidade é uma constante, o que não contribui para a posteriori condicional completa de \mathbf{a} .

A distribuição condicional completa para cada um dos efeitos aleatórios dos marcadores, segue

$$a_i | \boldsymbol{\beta}, \mathbf{a}_{-i}, \boldsymbol{\sigma}_a^2, \boldsymbol{\sigma}_e^2, \mathbf{y} \sim N \left[(\mathbf{z}_i' \mathbf{V}_{k-1}^{-1} \mathbf{z}_i + (\boldsymbol{\sigma}_{a_i}^2)^{-1})^{-1} \mathbf{z}_i' \mathbf{V}_{k-1}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}), [\mathbf{z}_i' \mathbf{V}_{k-1}^{-1} \mathbf{z}_i + (\boldsymbol{\sigma}_{a_i}^2)^{-1}]^{-1} \right] \tag{3.17}$$

A condicional completa para efeito aleatório a_i apresenta como solução

$$\begin{aligned}
\hat{a}_i &= (\mathbf{z}'_i \mathbf{V}_{k-1}^{-1} \mathbf{z}_i + (\sigma_{a_i}^2)^{-1})^{-1} \mathbf{z}'_i \mathbf{V}_{k-1}^{-1} (\mathbf{y} - \mathbf{X} \hat{\boldsymbol{\beta}}) \\
&= \sigma_{a_i}^2 \mathbf{z}'_i (\mathbf{V}_{k-1} + \mathbf{z}_i \sigma_{a_i}^2 \mathbf{z}'_i)^{-1} (\mathbf{y} - \mathbf{X} \hat{\boldsymbol{\beta}}) && \text{com } \mathbf{V}_{k-1} = \mathbf{V} - \mathbf{z}_i \mathbf{z}'_i \sigma_{a_i}^2 \\
&= \sigma_{a_i}^2 \mathbf{z}'_i \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X} \hat{\boldsymbol{\beta}}) && \text{com } \mathbf{V} = \mathbf{Z} \mathbf{Z}' \sigma_{a_i}^2 + \mathbf{I} \sigma_e^2
\end{aligned} \tag{3.18}$$

Observa-se que, ao comparar a solução da GWSAS com a solução dada pelo método GWS, como segue

$$\begin{aligned}
\hat{\mathbf{a}}_{GWS} &= (\mathbf{Z}' \mathbf{R}^{-1} \mathbf{Z} + \mathbf{G}^{-1})^{-1} \mathbf{Z}' \mathbf{R}^{-1} (\mathbf{y} - \mathbf{X} \hat{\boldsymbol{\beta}}) \\
&= \mathbf{G} \mathbf{Z}' (\mathbf{Z} \mathbf{G} \mathbf{Z}' + \mathbf{R})^{-1} (\mathbf{y} - \mathbf{X} \hat{\boldsymbol{\beta}})
\end{aligned} \tag{3.19}$$

uma implicação decorrente da solução apresentada em 3.18 em comparação com a solução apresentada pelo método de GWS 3.19 supondo $\mathbf{R} = \mathbf{I} \sigma_e^2$ e \mathbf{G} , é percebida pela correção da matriz \mathbf{V} , que remove $\mathbf{z}_i \mathbf{z}'_i \sigma_{a_i}^2$ referente ao marcador i e passa a ser fundamental para consistência do modelo GWSAS.

Observando-se as expressões 3.16 e 3.17, nota-se que todas as condicionais completas têm densidades conhecidas e, portanto, conduziu-se a obtenção das amostradas, usando técnicas MCMC, algoritmo amostrador de Gibbs, permitindo, assim, inferir sobre os parâmetros considerados no modelo.

O problema a destacar associou-se ao fato de que, para cada locus, várias rodadas do cálculo de inversão de \mathbf{V}_{k-1}^{-1} , são necessárias. Uma solução, para superar o problema da inversão da matriz \mathbf{V}_{k-1}^{-1} , foi o uso de técnicas de mudança de variáveis e produto de vetores para elevar a facilidade computacional.

3.2.4.1 Inversão da Matriz \mathbf{V}_{k-1}

Nesta proposta, à matriz \mathbf{V}_{k-1} competiu substituir por $\mathbf{V}_{k-1} = \mathbf{V} - \mathbf{z}_i \mathbf{z}'_i \sigma_{a_i}^2$ e em seguida, recorreu-se à multiplicação da matriz \mathbf{V}_{k-1}^{-1} pelos vetores \mathbf{z}'_i à esquerda e \mathbf{z}_i e à direita, simultaneamente. O mesmo foi realizado para a multiplicação de \mathbf{V}_{k-1}^{-1} pelos vetores \mathbf{z}'_i e $\tilde{\mathbf{y}}$, à esquerda e à direita, respectivamente. De modo consequente

$$\mathbf{z}'_i \mathbf{V}_{k-1}^{-1} \mathbf{z}_i = \mathbf{z}'_i [\mathbf{V} - \mathbf{z}_i \mathbf{z}'_i \sigma_{a_i}^2]^{-1} \mathbf{z}_i = \mathbf{z}'_i \mathbf{V}^{-1} \mathbf{z}_i - \mathbf{z}'_i \mathbf{V}^{-1} \mathbf{z}_i [\mathbf{z}'_i \mathbf{V}^{-1} \mathbf{z}_i - (\sigma_{a_i}^2)^{-1}]^{-1} \mathbf{z}'_i \mathbf{V}^{-1} \mathbf{z}_i \tag{3.20}$$

$$\mathbf{z}'_i \mathbf{V}_{k-1}^{-1} \tilde{\mathbf{y}} = \mathbf{z}'_i [\mathbf{V} - \mathbf{z}_i \mathbf{z}'_i \sigma_{a_i}^2]^{-1} \tilde{\mathbf{y}} = \mathbf{z}'_i \mathbf{V}^{-1} \tilde{\mathbf{y}} - \mathbf{z}'_i \mathbf{V}^{-1} \mathbf{z}_i [\mathbf{z}'_i \mathbf{V}^{-1} \mathbf{z}_i - (\sigma_{a_i}^2)^{-1}]^{-1} \mathbf{z}'_i \mathbf{V}^{-1} \tilde{\mathbf{y}} \tag{3.21}$$

e pela situação, a matriz inversa $[\mathbf{V} - \mathbf{z}_i \mathbf{z}_i' \sigma_{a_i}^2]^{-1}$ relacionou-se ao método de soma de matriz inversa, baseado na fórmula de Sherman e Morrison (1950). A solução requereu a geração de outra mudança de variável, tal $\phi = \mathbf{z}_i' \mathbf{V}_{k-1}^{-1} \mathbf{z}_i$ e $\omega = \mathbf{z}_i' \mathbf{V}_{k-1}^{-1} \tilde{\mathbf{y}}$. Assim sendo

$$\mathbf{z}_i' \mathbf{V}_{k-1}^{-1} \mathbf{z}_i = \phi - \phi [\phi - (\sigma_{a_i}^2)^{-1}]^{-1} \phi \quad (3.22)$$

$$\mathbf{z}_i' \mathbf{V}_{k-1}^{-1} \tilde{\mathbf{y}} = \omega - \phi [\phi - (\sigma_{a_i}^2)^{-1}]^{-1} \omega \quad (3.23)$$

o que permitiu compor a função distribuição condicional completa a posteriori de (3.17) como

$$a_i | \boldsymbol{\beta}, \mathbf{a}_{-i}, \sigma_a^2, \sigma_e^2, \mathbf{y} \sim N \left[(\phi - \phi \xi^{-1} + (\sigma_{a_i}^2)^{-1})^{-1} (\omega - \phi \xi^{-1} \omega), (\phi - \phi \xi^{-1} + (\sigma_{a_i}^2)^{-1})^{-1} \right] \quad (3.24)$$

em que $\xi^{-1} = [\phi - (\sigma_{a_i}^2)^{-1}]^{-1}$

3.2.5 Métodos de Análise para Seleção e Associação Genômica Ampla

Cada um dos métodos discutidos, a seguir, diferiram em relação a priori utilizada para o vetor de efeito do marcador \mathbf{a} . Declararam-se os métodos como RR-BLUP modificado, Bayes A modificado, Bayes B modificado e BSSV modificado. Os métodos de seleção genômica GWS, utilizados para comparações são RR-BLUP, Bayes A, Bayes B, propostos por Meuwissen, Hayes e Goddard (2001) e o BSSV proposto por Yi, George e Allison (2003). Os conceitos e definições de cada um dos métodos foram apresentados no referencial teórico. Os métodos RR-BLUP, Bayes A e Bayes B foram obtidos como uso dos pacotes rrBLUP e BGLR para prever o efeito dos marcadores e valores genômicos.

Os métodos de associação genômica GWAS, para comparações e verificação de eficácia pelos métodos modificados foi a função GWAS implementada pelo pacote rrBLUP no software R, executada a análise de associação genômica com base no modelo de Yu et al. (2006).

Para verificar a eficácia dos métodos, foram propostas análises para comparação de todos os métodos GWSAS com os chamados algoritmos FASTmrMLM e FASTmrEMMA (Wang et al. (2016); Wen et al. (2017)) dos estudos de associação de genômica multiloci, para verificar a capacidade de detectar genes previamente simulados e reais

3.2.5.1 RR-BLUP Modificado

Para o RR-BLUP modificado, requereu-se a especificação do vetor de efeitos fixos $\boldsymbol{\beta}$ e o vetor de efeitos aleatórios \mathbf{a} , conforme modelo 3.2.

1. De acordo com a distribuição a posteriori condicional completa normal, dada pela equação 3.16, a solução para os coeficientes de efeito fixo

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \quad (3.25)$$

2. Para RR-BLUP modificado, a distribuição a priori, para os efeitos aleatórios, considerando as variâncias iguais para $i = 1, \dots, k$, ou seja, $\sigma_{a_1}^2 = \dots = \sigma_{a_k}^2$, correspondeu à mesma identificação da priori para o RR-BLUP (MEUWISSEN; HAYES; GODDARD, 2001). Assim, a solução para os efeitos aleatórios, dada pelo núcleo da equação 3.17

$$\begin{aligned} \hat{a}_i &= (\mathbf{z}'_i\mathbf{V}_{k-1}^{-1}\mathbf{z}_i + (\sigma_a^2)^{-1})^{-1}\mathbf{z}'_i\mathbf{V}_{k-1}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \\ &= \sigma_a^2\mathbf{z}'_i(\mathbf{V}_{k-1} + \mathbf{z}_i\sigma_a^2\mathbf{z}'_i)^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) && \text{com } \mathbf{V}_{k-1} = \mathbf{V} - \mathbf{z}_i\mathbf{z}'_i\sigma_a^2 \\ &= \sigma_a^2\mathbf{z}'_i\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) && \text{com } \mathbf{V} = \mathbf{Z}\mathbf{Z}'\sigma_a^2 + \mathbf{I}\sigma_e^2 \\ &= \mathbf{z}'_i\left(\mathbf{Z}\mathbf{Z}' + \frac{\sigma_e^2}{\sigma_a^2}\right)^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) && \text{para } \sigma_a^2 > 0 \\ &= \mathbf{z}'_i(\mathbf{Z}\mathbf{Z}' + \lambda\mathbf{I})^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \end{aligned} \quad (3.26)$$

com variância é dada por

$$\text{Var}(a_i) = [\mathbf{z}'_i\mathbf{V}_{k-1}^{-1}\mathbf{z}_i + (\sigma_a^2)^{-1}]^{-1}$$

Como o amostrador de Gibbs utilizado para delinear as incógnitas, conceituou-se o processo de inversão da matriz \mathbf{V}_{k-1} , prosseguindo com a distribuição a posteriori condicional completa para amostras de a_i

$$a_i|\boldsymbol{\beta}, \mathbf{a}_{-i}, \sigma_a^2, \sigma_e^2, \mathbf{y} \sim N\left[(\phi - \phi\xi^{-1} + (\sigma_{a_i}^2)^{-1})^{-1}(\omega - \phi\xi^{-1}\omega), (\phi - \phi\xi^{-1} + (\sigma_{a_i}^2)^{-1})^{-1}\right] \quad (3.27)$$

tal que $\phi = \mathbf{z}'_i\mathbf{V}_{k-1}^{-1}\mathbf{z}_i$, $\omega = \mathbf{z}'_i\mathbf{V}_{k-1}^{-1}\tilde{\mathbf{y}}$ e $\xi^{-1} = [\phi - (\sigma_{a_i}^2)^{-1}]^{-1}$.

3. A distribuição condicional completa para amostrar o parâmetro de variância σ_a^2

$$p(\sigma_a^2|\boldsymbol{\beta}, \mathbf{a}, \sigma_e^2, \mathbf{y}) \propto (\sigma_a^2)^{-\left(\frac{k}{2} + \frac{v_a}{2} + 1\right)} e^{-\frac{1}{2\sigma_a^2}(\mathbf{a}'\mathbf{a} + v_a s_a^2)} \quad (3.28)$$

reconhecida como o núcleo de uma qui-quadrado escalonada invertida $\sigma_a^2|\boldsymbol{\beta}, \mathbf{a}, \sigma_e^2, \mathbf{y} \sim \chi^{-2}\left(v_a + k, \frac{\mathbf{a}'\mathbf{a} + v_a s_a^2}{v_a + k}\right)$, com $v_a + k$ graus de liberdade e parâmetro escala $\frac{\mathbf{a}'\mathbf{a} + v_a s_a^2}{v_a + k}$. A

amostra, para esta condicional a posteriori completa obteve-se de $\frac{(\mathbf{a}'\mathbf{a} + v_a s_a^2)}{\chi_{(v_a+k)}^2}$, na qual $\chi_{(v_a+k)}^2$ admite-se uma qui-quadrado com $v_a + k$ graus de liberdade.

4. A condicional completa para σ_e^2

$$p(\sigma_e^2 | \boldsymbol{\beta}, \mathbf{a}, \sigma_a^2, \mathbf{y}) \propto (\sigma_e^2)^{-\left(\frac{n}{2} + \frac{v_e}{2} + 1\right)} e^{-\frac{1}{2\sigma_e^2}(\mathbf{e}'\mathbf{e} + v_e s_e^2)} \quad (3.29)$$

que pode ser reconhecida como o núcleo de uma qui-quadrado escalonada invertida $\sigma_e^2 | \boldsymbol{\beta}, \mathbf{a}, \sigma_a^2, \mathbf{y} \sim \chi^{-2}\left(v_e + n, \frac{\mathbf{e}'\mathbf{e} + v_e s_e^2}{v_e + n}\right)$, com $v_e + n$ graus de liberdade e parâmetro escala $\frac{\mathbf{e}'\mathbf{e} + v_e s_e^2}{(v_e + n)}$. A amostra para esta condicional a posteriori completa é obtida de $\frac{(\mathbf{e}'\mathbf{e} + v_e s_e^2)}{\chi_{(v_e+n)}^2}$, onde $\chi_{(v_e+n)}^2$ é uma qui-quadrado com $v_e + n$ graus de liberdade.

3.2.5.2 Bayes A Modificado

Como mencionado, o método de Bayes A proposto por Meuwissen, Hayes e Goddard (2001), difere do RR-BLUP, em que cada SNP tem sua própria variância. No estudo atual, atribui-se ao método Bayes A modificado, considerando suas distribuições condicionais como segue.

A priori usada para $\boldsymbol{\beta}$ foi a flat, considerada em todos os modelos. Assim, a condicional completa para $\boldsymbol{\beta}$ passou a ser idêntica ao método RR-BLUP modificado (ver Equação 3.16).

1. A condicional completa, para efeito aleatório a_i sob o modelo Bayes A, segue conforme detalhada anteriormente em 3.17, com solução

$$\begin{aligned} \hat{a}_i &= (\mathbf{z}_i' \mathbf{V}_{k-1}^{-1} \mathbf{z}_i + (\sigma_{a_i}^2)^{-1})^{-1} \mathbf{z}_i' \mathbf{V}_{k-1}^{-1} (\mathbf{y} - \mathbf{X} \hat{\boldsymbol{\beta}}) \\ &= \sigma_{a_i}^2 \mathbf{z}_i' (\mathbf{V}_{k-1} + \mathbf{z}_i \sigma_{a_i}^2 \mathbf{z}_i')^{-1} (\mathbf{y} - \mathbf{X} \hat{\boldsymbol{\beta}}) && \text{com } \mathbf{V}_{k-1} = \mathbf{V} - \mathbf{z}_i \mathbf{z}_i' \sigma_{a_i}^2 \\ &= \sigma_{a_i}^2 \mathbf{z}_i' \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X} \hat{\boldsymbol{\beta}}) && \text{com } \mathbf{V} = \mathbf{Z} \mathbf{Z}' \sigma_{a_i}^2 + \mathbf{I} \sigma_e^2 \end{aligned} \quad (3.30)$$

com variância dada por

$$\text{Var}(a_i) = [\mathbf{z}_i' \mathbf{V}_{k-1}^{-1} \mathbf{z}_i + (\sigma_{a_i}^2)^{-1}]^{-1}$$

2. A distribuição condicional completa para o parâmetro de variância $\sigma_{a_i}^2$

$$p(\sigma_{a_i}^2 | \boldsymbol{\beta}, \mathbf{a}, \sigma_e^2, \mathbf{y}) \propto (\sigma_{a_i}^2)^{-\left(\frac{k}{2} + \frac{v_a}{2} + 1\right)} e^{-\frac{1}{2\sigma_{a_i}^2}(\mathbf{a}'\mathbf{a} + v_a s_a^2)} \quad (3.31)$$

reconhecida como o núcleo de uma qui-quadrado escalonada invertida $\sigma_{a_i}^2 | \boldsymbol{\beta}, \mathbf{a}, \sigma_e^2, \mathbf{y} \sim \chi^{-2} \left(v_a + k, \frac{\mathbf{a}'\mathbf{a} + v_a s_a^2}{v_a + k} \right)$, com $v_a + k$ graus de liberdade e parâmetro escala $\frac{\mathbf{a}'\mathbf{a} + v_a s_a^2}{(v_a + k)}$. A amostra para esta condicional a posteriori completa obteve-se de $\frac{(\mathbf{a}'\mathbf{a} + v_a s_a^2)}{\chi^2_{(v_a + k)}}$, na qual $\chi^2_{(v_a + k)}$ se admite uma qui-quadrado com $v_a + k$ graus de liberdade.

3. A priori para σ_e^2 é a mesma em todos os modelos e a distribuição condicional completa para σ_e^2 , idêntica à definida para o método RR-BLUP modificado (Equação 3.29).

3.2.5.3 Bayes B Modificado

No método Bayes B proposto por Meuwissen, Hayes e Goddard (2001), assumiu-se que muitos dos marcadores não teriam efeito, ou seja, considerou-se uma distribuição de mistura, em que cada marcador tem ponto de massa zero com probabilidade π e uma distribuição t-univariada com probabilidade $(1 - \pi)$.

Para o método Bayes B modificado, as condicionais completas de $\boldsymbol{\beta}$ e σ_e^2 avançaram como as mesmas que do RR-BLUP modificado. O interessante para o método Bayes B modificado foi o efeito dos marcadores que se escreveu com uma proporção de mistura que se tornou uma variável aleatória de Bernoulli. A variável δ_i para o locus i assumiu valor 1, teve probabilidade $(1 - \pi)$ e, quando assumiu valor zero, teve probabilidade π .

$$p(\delta_i) = (1 - \pi)^{\delta_i} + \pi^{\delta_i}$$

Quando $\delta_i = 1$, a distribuição a posteriori condicional completa, para efeito dos marcadores a_i adotou-se idêntica à utilizada pelo método Bayes A modificado.

A agregação numérica das distribuições a posteriori condicionais completas encarregou-se ao algoritmo de Monte Carlo da cadeia de Markov, por meio da amostragem de Gibbs (CASSELLA; GEORGE, 1992), descrito pelas etapas:

1. Para o vetor $\boldsymbol{\beta}$ de efeitos fixos, as amostras seguiram a distribuição a posteriori condicional completa

$$\boldsymbol{\beta} | \mathbf{a}, \sigma_a^2, \sigma_e^2, \mathbf{y} \sim N \left[(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \mathbf{X}'\mathbf{V}^{-1}(\mathbf{y}), [\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}]^{-1} \right] \quad (3.32)$$

2. As amostras de a_i foram geradas a partir da distribuição a posteriori condicional completa

$$a_i | \boldsymbol{\beta}, \mathbf{a}_{-i}, \boldsymbol{\sigma}_a^2, \boldsymbol{\sigma}_e^2, \mathbf{y} \sim N \left[(\phi - \phi \xi^{-1} + (\boldsymbol{\sigma}_{a_i}^2)^{-1})^{-1} (\omega - \phi \xi^{-1} \omega), (\phi - \phi \xi^{-1} + (\boldsymbol{\sigma}_{a_i}^2)^{-1})^{-1} \right] \quad (3.33)$$

tal que a variância definiu-se $v_{a_i} = \delta_i \boldsymbol{\sigma}_{a_i}^2 + (1 - \delta_i) \varepsilon$, com $\delta_i \in 0, 1$ e $p(\delta_i) \sim \text{Bernoulli}(\pi)$. A variável $\varepsilon = 10^{-6}$, a priori e correspondente à variação individual do marcador, a 0,5% da variação fenotípica, isto é, $\varepsilon = \frac{\sigma_y^2 \times 0,005}{k}$.

3. As amostras de $\boldsymbol{\sigma}_{a_i}^2$ geradas da distribuição qui-quadrado escalonada invertida com os parâmetros

$$p(\boldsymbol{\sigma}_{a_i}^2 | \boldsymbol{\beta}, \mathbf{a}, \boldsymbol{\sigma}_e^2, \mathbf{y}) \propto (\boldsymbol{\sigma}_{a_i}^2)^{-\left(\frac{1}{2} + \frac{v_a}{2} + 1\right)} e^{-\frac{1}{2\boldsymbol{\sigma}_{a_i}^2} (a_i^2 + v_a s_a^2)} \quad (3.34)$$

com $v_a + 1$ graus de liberdade e parâmetro escala $\frac{a_i^2 + v_a s_a^2}{v_a + 1}$.

4. Amostra de δ_i obtida da distribuição a posteriori condicional completa

$$p(\delta_i | \boldsymbol{\beta}, \mathbf{a}, \boldsymbol{\sigma}_e^2, \mathbf{y}) = \frac{(1 - \pi) N(a_i | 0, \boldsymbol{\sigma}_{a_i}^2)}{\pi N(a_i | 0, \boldsymbol{\sigma}_{a_i}^2) + (1 - \pi) N(a_j | 0, \varepsilon)}$$

5. A priori para $\boldsymbol{\sigma}_e^2$ foi a mesma, em todos os modelos e a distribuição condicional completa para $\boldsymbol{\sigma}_e^2$, idêntica à definida para o método RR-BLUP modificado (Equação 3.29).

3.2.5.4 BSSV Modificado

O método BSSV modificado, sofreu ajustes a partir do modelo original proposto por Yi, George e Allison (2003). O método bayesiano foi utilizado pela capacidade de selecionar marcadores de grande efeito em modelos de múltiplas marcas. Utilizou-se a nova abordagem, para abranger todos os efeitos de marcadores e o método de amostragem passa pelos passos:

1. Amostra de $\boldsymbol{\beta}$ da distribuição a posteriori condicional completa

$$\boldsymbol{\beta} | \mathbf{a}, \boldsymbol{\sigma}_a^2, \boldsymbol{\sigma}_e^2, \mathbf{y} \sim N \left[(\mathbf{X}' \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}' \mathbf{V}^{-1} (\mathbf{y}), [\mathbf{X}' \mathbf{V}^{-1} \mathbf{X}]^{-1} \right] \quad (3.35)$$

2. As amostras de a_i foram geradas a partir da distribuição a posteriori condicional completa

$$a_i | \boldsymbol{\beta}, \mathbf{a}_{-i}, \boldsymbol{\sigma}_a^2, \boldsymbol{\sigma}_e^2, \mathbf{y} \sim N \left[(\phi - \phi \xi^{-1} + (\boldsymbol{\sigma}_{a_i}^2)^{-1})^{-1} (\omega - \phi \xi^{-1} \omega), (\phi - \phi \xi^{-1} + (\boldsymbol{\sigma}_{a_i}^2)^{-1})^{-1} \right] \quad (3.36)$$

tal que a variância define-se $v_{a_i} = \delta_i \boldsymbol{\sigma}_{a_i}^2 + (1 - \delta_i) \varepsilon$, com $\delta_i \in 0, 1$ e $p(\delta_i) \sim \text{Bernoulli}(\pi)$. A variável $\varepsilon = 10^{-6}$, a priori e correspondente à variação individual do marcador, a 0,5% da variação fenotípica, isto é, $\varepsilon = \frac{\sigma_y^2 \times 0,005}{k}$.

3. As amostras de $\boldsymbol{\sigma}_{a_i}^2$ foram da distribuição qui-quadrado escalonada invertida com os parâmetros:

$$p(\boldsymbol{\sigma}_{a_i}^2 | \boldsymbol{\beta}, \mathbf{a}, \boldsymbol{\sigma}_e^2, \mathbf{y}) \propto (\boldsymbol{\sigma}_{a_i}^2)^{-\left(\frac{1}{2} + \frac{v_a}{2} + 1\right)} e^{-\frac{1}{2\boldsymbol{\sigma}_{a_i}^2} (a_i^2 + v_a s_a^2)} \quad (3.37)$$

com $v_a + 1$ graus de liberdade e parâmetro escala. $\frac{a_i^2 + v_a s_a^2}{v_a + 1}$.

4. Amostra de δ_i obtida da distribuição a posteriori condicional completa

$$p(\delta_i | \boldsymbol{\beta}, \mathbf{a}, \boldsymbol{\sigma}_e^2, \mathbf{y}) = \frac{(1 - \pi) N(a_i | 0, \boldsymbol{\sigma}_{a_i}^2)}{\pi N(a_i | 0, \boldsymbol{\sigma}_{a_i}^2) + (1 - \pi) N(a_j | 0, \varepsilon)}$$

5. Amostra de π_i foi obtida de uma distribuição condicional:

$$p(\pi | \delta_j) = \text{Beta} \left[1 + \sum_{j=1}^k \delta_j, 1 + \left(k - \sum_{j=1}^k \delta_j \right) \right] \quad (3.38)$$

6. A priori para $\boldsymbol{\sigma}_e^2$ é a mesma, em todos os modelos e a distribuição condicional completa para $\boldsymbol{\sigma}_e^2$ é idêntica à definida para o método RR-BLUP modificado (Equação 3.29).

3.2.5.5 Implementação dos Métodos em R

Para implementar os métodos apresentados nas subseções anteriores, empregou-se a linguagem R. Os *scripts* são demonstrados no Anexo para fins ilustrativos. Em primeiro lugar, os genótipos dos marcadores e os fenótipos das características foram utilizados para estimar os efeitos do marcadores.

De posse dos dados fenotípicos [oligogênico herdabilidade 0,2 (yo2), oligogênico herdabilidade 0,5 (yo5), oligogênico herdabilidade 0,8 (yo8), poligênico herdabilidade 0,2 (yp2),

poligênico herdabilidade 0,5 (yp5) e poligênico herdabilidade 0,8 (yp8)], obtidos por simulação e da matriz genotípica das três populações (matriz \mathbf{W}), nos tópicos seguintes foi detalhada a implementação por métodos computacionais (utilizando *software* R e ambiente edição RStudio (RSTUDIO, 2016)). Aplicou-se a dinâmica de amostragem Método Monte Carlo, via Cadeia de Markov (GOLUB; VAN LOAN, 2012), dentro do amostrador de Gibbs (CASELLA; GEORGE, 1992), para cada distribuição a posteriori condicional completa, afim de alcançar informações sobre os parâmetros dos modelos.

Os elementos do algoritmo Gibbs colhem as amostras dos efeitos dos marcadores a_i , definindo os passos:

- i) Cadeia inicial gerada com 3000 iterações, para todos os parâmetros.
- ii) Descarte dos 300 primeiros valores (processo *burn-in*, elimina problemas de efeito do valor inicial e dependência dos parâmetros).
- iii) Pontos amostrais tomados de 3 em 3 iterações, de forma obter amostra aleatória (técnica *jump*).

Processou-se os passos descritos acima, para cada método, até que a convergência fosse atingida. As amostras finais de cada cadeia constaram de 900 observações. O número de iterações 3000 foi definido pelo mesmo padrão do pacote GAPIT. A quantidade de 6000 marcadores do tipo SNP torna o processo de análise intratável pelos critérios de convergência em análise Bayesiana, mas o fato de determinar soluções viáveis e equiparáveis aos métodos GWS (ver Gráficos 4.4), podem-se considerar os resultados.

3.2.6 Teste de Wald

Das amostras resultantes do processo de Gibbs, as inferências avançaram quanto ao efeito dos marcadores. O efeito de significância dos marcadores foi determinado pelo teste Wald, tendo em vista os resultados das cadeias. A razão estatística, resultante deste teste W_i , $i = 1, \dots, k$ sob a hipótese $H_0 : a_i = 0$, segue distribuição normal padrão. Os valores de teste foram obtidos por

$$W_i = \frac{a_i^2}{\sigma_{a_i}^2} \quad (3.39)$$

O valor-p é definido como

$$\text{valor-p} = p[|Z| > |W_i| | H_0] = 2p[Z > |W_i| | H_0] \quad (3.40)$$

sendo que Z denota a variável aleatória da distribuição normal padrão. O nível de significância mínimo para uma marca apresentar valor diferente de zero, foi dado como de 5%.

Os exames de diagnóstico por imagens gráficas foram analisados detalhadamente tanto em relação aos aspectos do diagnóstico dos métodos de seleção GWS, como dos métodos modificados em comparação com a técnica GWAS. Procurou-se caracterizar o encolhimento das estimativas dos efeitos, as quantidades de falsos positivos, os falsos negativos.

Os gráficos Manhattan, de soluções dos efeitos dos SNPs foram obtidos em todos os fenótipos, ou seja, cada vetor de observação fenotípica foi criado nos cenários oligogênico e poligênico com herdabilidade $h^2 \in \{0, 2; 0, 5; 0, 8\}$. O eixo Y representa o efeito ou $-\log_{10}(\text{valor} - p)$, ou o $\text{valor} - p$ do teste de Wald (Equação 3.39 e 3.40), e o eixo X representa o SNP, dentro de seu respectivo cromossomo.

3.3 Critérios de Comparação

A Soma de Quadrados de valores Residuais Predito (*Predicted Residual Sum of Squares*-PRESS) é uma estatística de validação de modelo usada como um medida de poder preditivo. Também pode ser usada para comparar modelos de regressão. Para um conjunto de dados de tamanho n , PRESS é dado por

$$\text{PRESS} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_{i(-i)})^2.$$

calculado para a observação omitida em cada modelo de regressão fixado. Em que $\hat{y}_{i(-i)}$ é o valor estimado de $n - 1$ observações restantes. Em geral, quanto menor o valor da PRESS, melhor será o poder preditivo do modelo. A PRESS pode ser usada para calcular o valor de R^2

$$R_{pred}^2 = 1 - \frac{\text{PRESS}}{\text{SST}}$$

em que $\text{SST} = \sum_{i=1}^n (y_i - \bar{y})^2$ é a Soma de Quadrados Total (*Total Sum of Squares*-SST), com \bar{y} , a média da variável y . O R^2 sempre assume um valor entre 0 e 1. Quanto mais próximo o R^2 é

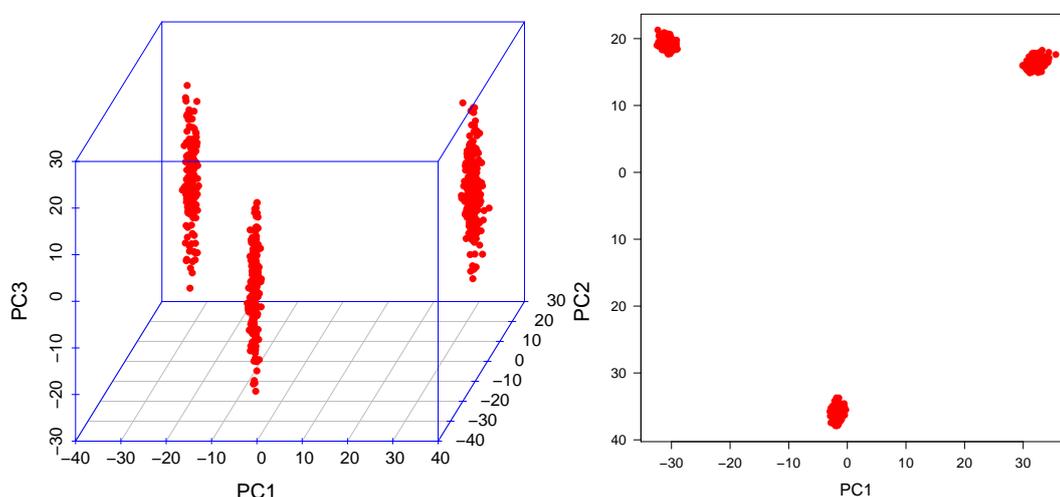
de um, melhor a adequação do modelo. Estas estatísticas foram assim calculadas e tornam-se úteis para validar a capacidade preditiva de seu modelo.

4 RESULTADOS E DISCUSSÃO

4.1 Diagnóstico Genotípico

Os resultados em estudos de GWAS consideram-se influenciados pela estrutura genética da população, que fornece informações sobre a ancestralidade genética de indivíduos, história evolutiva da população, padrões de dispersão da população (WANG et al., 2014). Para que os métodos propostos capturassem de forma eficiente a informação sobre a estrutura da população, a matriz de marcadores SNPs foi simulada sendo representativa de três populações distintas.

Figura 4.1 – Gráfico de dispersão dos três primeiros componentes principais mostra a distribuição entre as três populações simuladas



Fonte: Da autora (2019)

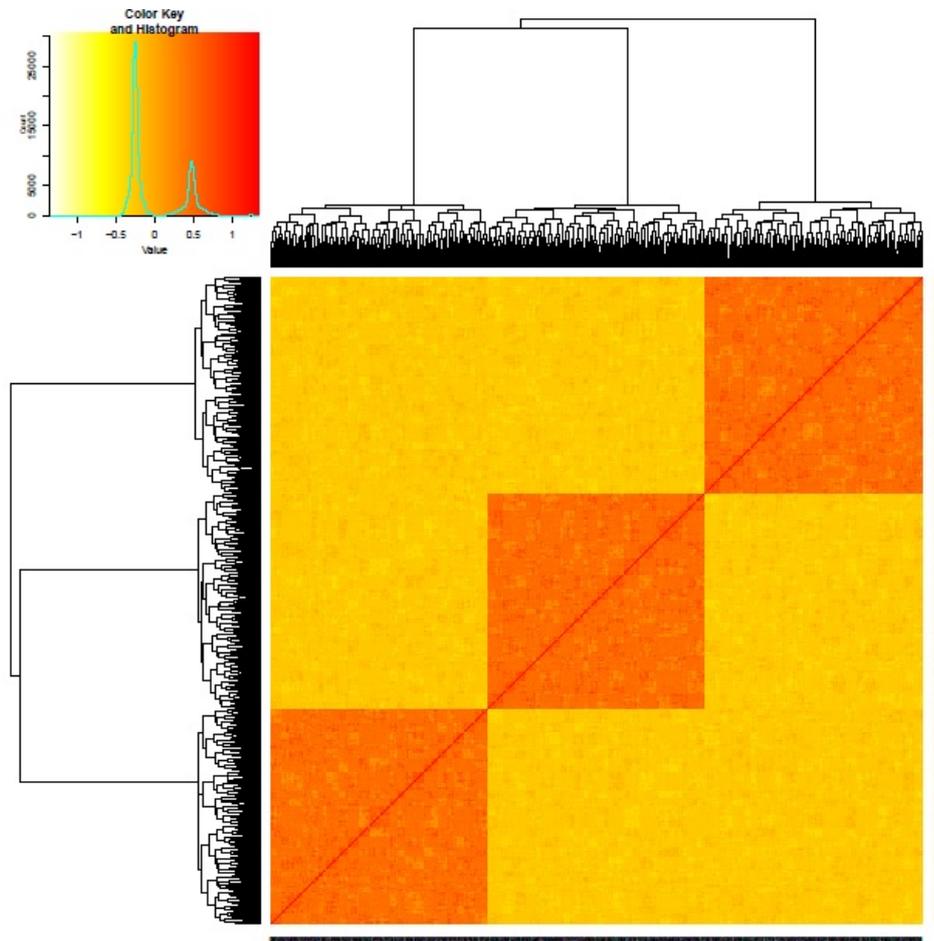
Ao avaliar a exigência do pressuposto de distintas populações, realizou-se a análise gráfica de componentes principais (*Principal Component Analysis* - PCA) que resume a variação genética por região dos dados simulados. Observaram pela Figura 4.1 os primeiros três eixos do PCA, com um agrupamento claro entre os eixos, indicando um alto nível de estrutura populacional, condizente com as três formações de população, simulados os genótipos.

A visualização dos dados simulados, na Figura 4.2, tanto pelo dendograma como pelo mapa de calor (*Heatmap*), permitiu reconhecer três blocos alaranjados, o que indica clareza expressiva dos SNPs e delimita a formação nítida de três diferentes grupos populacionais.

4.1.1 Análise do Desequilíbrio de Ligação

Outra informação para os estudos GWAS é o interesse em nível de desequilíbrio de ligação (*Linkage Disequilibrium*-LD) presente na população. Conforme o desequilíbrio de ligação,

Figura 4.2 – *Heatmap* e dendrograma da matriz de parentesco mostrando nos dados simulados. Vermelho indica alta correlação entre os pares de marcadores e o amarelo baixa correlação

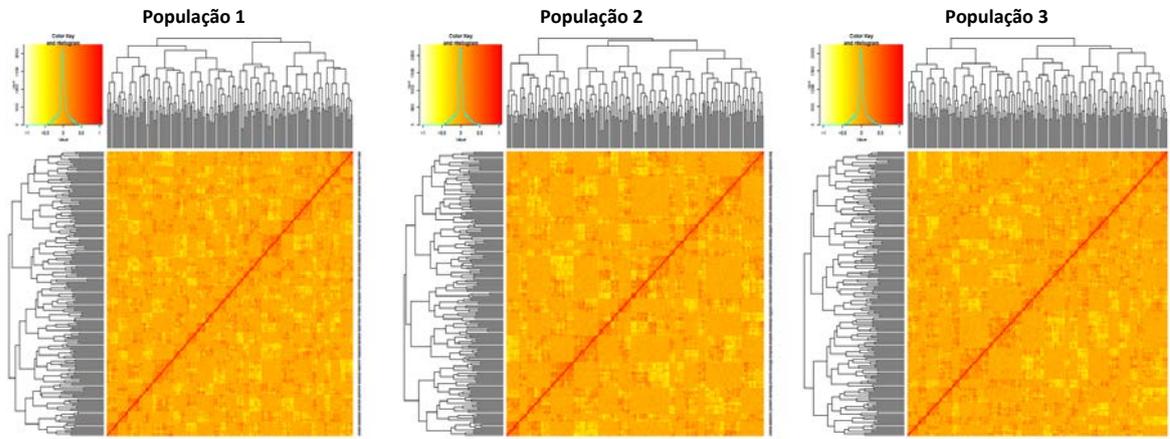


Fonte: Da autora (2019)

decai em função da distância, ao longo do cromossomo, interfere no número de marcadores ao acaso, necessários para encontrar associações significativas.

Alternativamente, avaliou-se o desequilíbrio de ligação, por meio das matrizes de desequilíbrio (Figura 4.3), visualizando o arranjo linear do desequilíbrio de ligação entre loci ao longo do cromossomo. A matriz $\mathbf{W} = [w_{ij}]$, separadamente para as três distintas populações, ou seja, particionou-se a matriz \mathbf{W} em \mathbf{W}_1 , \mathbf{W}_2 e \mathbf{W}_3 (respectivas à População 1, População 2, População 3), cada uma com dimensão 200×6000 .

Figura 4.3 – Mapas de calor dos valores de desequilíbrio de ligação (LD) para genoma simulado em cada população. Os marcadores foram ordenados, nos eixos x e y, de modo que cada célula do mapa de calor representa um único par de marcadores.



Fonte: Da autora (2019)

A construção dos gráfico de LD, as estimativas da estrutura da população e matrizes de PCA realizaram-se com auxílio do *software* computacional R-3.4.3 (64 bit) (R DEVELOPMENT CORE TEAM, 2016), por meio do pacote GAPIT (*Genome Association and Prediction Integrated Tool*) (LIPKA et al., 2012).

4.1.2 Diagnóstico Fenotípico

Para a simulação dos fenótipos, considerou-se o erro de uma distribuição gaussiana com variância $\sigma_e^2 = \sigma_a^2 \left(\frac{1-h^2}{h^2} \right)$, dependendo da herdabilidade $h^2 \in \{0,2;0,5;0,8\}$ os quais σ_a^2 e σ_e^2 são componentes da variância genética aditiva e residual, respectivamente (Tabela 4.1).

Tabela 4.1 – Estimativas da variância genética aditiva (σ_a^2) e residual (σ_e^2) para cada vetor fenotípico simulado, considerando herdabilidades $h^2 \in \{0,2;0,5;0,8\}$

| Cenários | h^2 | σ_a^2 | σ_e^2 | $\sigma_a^2 + \sigma_e^2$ |
|--------------------|-------|--------------|--------------|---------------------------|
| Oligogênico | 0,2 | 3,031 | 12,1256 | 16,1823 |
| | 0,5 | 3,031 | 3,03142 | 5,8171 |
| | 0,8 | 3,031 | 0,75785 | 3,7887 |
| Poligênico | 0,2 | 20,394 | 81,5789 | 102,4388 |
| | 0,5 | 20,394 | 20,394 | 40,3471 |
| | 0,8 | 20,394 | 5,09868 | 25,6868 |

Fonte: Da autora (2019)

Averiguou-se a eficiência dos valores de simulação, para caracterizar três distintas populações, por meio dos testes das médias ponderadas. Em cada população (I, II e III), de tamanho

$n_i = 200$, com i e $j \in \{1, 2, 3\}$, estimaram-se a média fenotípica (\bar{x}_i) e a variância fenotípica (s_i^2) observadas na população i . As estatísticas foram descritas pelas Tabelas 4.2 e 4.3.

Tabela 4.2 – Estatísticas das três populações distintas em cada cenário

| Modelo | População | n_i | \bar{x}_i | s_i^2 | $(\bar{x}_i - \bar{x}_j)$ | $t = \frac{(\bar{x}_i - \bar{x}_j)}{\sqrt{\frac{s_i^2}{n_i} + \frac{s_j^2}{n_j}}}$ |
|--------------------|-----------|-------|-------------|---------|----------------------------------|--|
| Oligogênico | I | 200 | -1,559 | 1,411 | $(\bar{x}_1 - \bar{x}_2)$ -2,891 | -23,667 |
| | II | 200 | 1,333 | 1,574 | $(\bar{x}_1 - \bar{x}_3)$ -1,784 | -13,949 |
| | III | 200 | 0,226 | 1,862 | $(\bar{x}_2 - \bar{x}_3)$ 1,107 | 8,444 |
| Poligênico | I | 200 | 2,227 | 14,349 | $(\bar{x}_1 - \bar{x}_2)$ 4,996 | 12,265 |
| | II | 200 | -2,769 | 18,834 | $(\bar{x}_1 - \bar{x}_3)$ 1,686 | 4,385 |
| | III | 200 | 0,541 | 15,230 | $(\bar{x}_2 - \bar{x}_3)$ -3,310 | -8,020 |

Fonte: Da autora (2019)

Tabela 4.3 – Estatísticas das três populações distintas após incorporação do resíduo para cada fenótipo

| Modelo | População | n_i | \bar{x}_i | s_i^2 | $(\bar{x}_i - \bar{x}_j)$ | $t = \frac{(\bar{x}_i - \bar{x}_j)}{\sqrt{\frac{s_i^2}{n_i} + \frac{s_j^2}{n_j}}}$ |
|--------|-----------|-------|-------------|---------|----------------------------------|--|
| yo2 | I | 200 | -1,732 | 15,992 | $(\bar{x}_1 - \bar{x}_2)$ -2,855 | -7,318 |
| | II | 200 | 1,123 | 14,456 | $(\bar{x}_1 - \bar{x}_3)$ -2,079 | -5,380 |
| | III | 200 | 0,347 | 13,880 | $(\bar{x}_2 - \bar{x}_3)$ 0,776 | 2,062 |
| yo5 | I | 200 | -1,488 | 4,568 | $(\bar{x}_1 - \bar{x}_2)$ -2,672 | -12,312 |
| | II | 200 | 1,184 | 4,851 | $(\bar{x}_1 - \bar{x}_3)$ -1,775 | -8,395 |
| | III | 200 | 0,287 | 4,372 | $(\bar{x}_2 - \bar{x}_3)$ 0,897 | 4,178 |
| yo8 | I | 200 | -1,534 | 1,924 | $(\bar{x}_1 - \bar{x}_2)$ -2,929 | -19,815 |
| | II | 200 | 1,394 | 2,446 | $(\bar{x}_1 - \bar{x}_3)$ -1,865 | -12,380 |
| | III | 200 | 0,331 | 2,615 | $(\bar{x}_2 - \bar{x}_3)$ 1,064 | 6,688 |
| yp2 | I | 200 | 1,951 | 104,666 | $(\bar{x}_1 - \bar{x}_2)$ 4,342 | 4,363 |
| | II | 200 | -2,391 | 93,455 | $(\bar{x}_1 - \bar{x}_3)$ 1,522 | 1,503 |
| | III | 200 | 0,428 | 100,466 | $(\bar{x}_2 - \bar{x}_3)$ -2,819 | -2,864 |
| yp5 | I | 200 | 2,388 | 37,953 | $(\bar{x}_1 - \bar{x}_2)$ 5,018 | 8,176 |
| | II | 200 | -2,629 | 37,376 | $(\bar{x}_1 - \bar{x}_3)$ 2,283 | 3,821 |
| | III | 200 | 0,106 | 33,432 | $(\bar{x}_2 - \bar{x}_3)$ -2,735 | -4,596 |
| yp8 | I | 200 | 2,323 | 20,252 | $(\bar{x}_1 - \bar{x}_2)$ 4,928 | 10,700 |
| | II | 200 | -2,605 | 22,175 | $(\bar{x}_1 - \bar{x}_3)$ 1,821 | 3,944 |
| | III | 200 | 0,501 | 22,409 | $(\bar{x}_2 - \bar{x}_3)$ -3,107 | -6,580 |

Fonte: Da autora (2019)

Fenótipos yo2: oligogênico herdabilidade 0,2; yo5: oligogênico herdabilidade 0,5; yo8: oligogênico herdabilidade 0,8; yp2: poligênico herdabilidade 0,2; yp5: poligênico herdabilidade 0,5 e yp8: poligênico herdabilidade 0,8

Assim, adotando $\alpha = 5\%$, a hipótese de que as médias fenotípicas de duas populações são iguais foi testada e a variável correspondente ao desvio rejeitada se $|t| > z_{\alpha/2}$ (a estatística

t-Student se aproxima da normal quando aumenta o número do grau de liberdade). Rejeitou-se a hipótese de igualdade entre as médias fenotípicas das populações para cada um dos modelos. Com base nestas amostras, pode-se afirmar, em nível de 5% de significância, que as populações diferem quanto ao caractere fenotípico, produzindo efeitos distintos para cada população. Destacou-se que, ao menos para as populações I e III, no modelo poligênico de herdabilidade 0,2, o nível de significância para rejeição passou a ser $\alpha = 10\%$.

Tabela 4.4 – Resumo dos resultados de seleção de modelo pelo critério de informação bayesiano (BIC) ajustado pelo pacote GAPIT

| Modelo | Número de Componentes | BIC | log da verossimilhança |
|--------|-----------------------|------------|------------------------|
| yo2 | 0 | -1661,7779 | -1652,1826 |
| | 1 | -1664,1500 | -1651,3561 |
| | 2 | -1667,3021 | -1651,3098 |
| | 3 | -1670,5434 | -1651,3526 |
| yo5 | 0 | -1278,2414 | -1268,6460 |
| | 1 | -1280,4858 | -1267,6919 |
| | 2 | -1283,6385 | -1267,6462 |
| | 3 | -1286,7505 | -1267,5597 |
| yo8 | 0 | -1012,5189 | -1002,9235 |
| | 1 | -1014,4493 | -1001,6555 |
| | 2 | -1017,5942 | -1001,6019 |
| | 3 | -1020,7678 | -1001,5770 |
| yp2 | 0 | -2229,0451 | -2219,4497 |
| | 1 | -2232,0226 | -2219,2288 |
| | 2 | -2235,2492 | -2219,2569 |
| | 3 | -2238,3481 | -2219,1573 |
| yp5 | 0 | -1885,4899 | -1875,8945 |
| | 1 | -1888,3328 | -1875,5389 |
| | 2 | -1891,5395 | -1875,5472 |
| | 3 | -1894,6280 | -1875,4372 |
| yp8 | 0 | -1642,3294 | -1632,7340 |
| | 1 | -1645,1934 | -1632,3995 |
| | 2 | -1648,3887 | -1632,3963 |
| | 3 | -1651,5283 | -1632,3375 |

Fonte: Da autora (2019)

A Tabela 4.4 destaca o número de componentes principais (PCs) incluídos nos modelos GWAS quando ajustados pelo pacote GAPIT. Observou-se que o valor de BIC (maior BIC, implica melhor número de componentes a considerar para o modelo final em GWAS) é conduzido para determinar o número ideal de PCs. Os valores das componentes principais ajustados para cada um dos fenótipos: yo2, yo5, yo8, yp2, yp5 e yp8 apontam zero compo-

mentos ($BIC = -1661,7779$, $BIC = -1278,2414$, $BIC = -1012,5189$, $BIC = -2229,0451$, $BIC = -1885,4899$ e $BIC = -1642,3294$, respectivamente). A determinação do número de componentes principais explica a estrutura de população, porém o que se destacou importante, nos resultados de saída fornecidos pelo GAPIT é que o número de PCS, a ser selecionado para inclusão no modelo final, não deteve o número real de PCs para cada fenótipo considerado. Neste caso, ao tentar explicar a estrutura de população, cautela no uso do critério BIC é relevante, uma vez que não corresponde, de forma determinada, a um número ideal de PCS/co-variáveis a serem incluídos para cada fenótipo, pelo modelo GWAS.

Lipka et al. (2012) afirmam que o grau de correlação com a estrutura populacional varia de característica para característica. Portanto o conjunto completo de PCs selecionados, para explicar a estrutura populacional no modelo GWAS, não é necessário para todas as características. Como tal, O GAPIT tem a capacidade de conduzir a seleção de modelos, baseada no critério de informação Bayesiano (BIC), para o número ideal de PCs para inclusão nos modelos GWAS.

4.1.3 Análise Comparativa dos Métodos

Ao avaliar o desempenho de cada método proposto, foram diagnosticadas informações dos valores de $-\log_{10}(\text{valor} - p)$, $\text{valor} - p$ do teste de Wald, efeitos dos 6000 marcadores do tipo SNPs, contidas nos gráficos Manhattan, para cada vetor de observação fenotípica criado nos cenários oligogênico e poligênico com herdabilidade $h^2 \in \{0,2;0,5;0,8\}$. As Figuras 4.4, 4.5, 4.6, 4.7, 4.8, 4.9, 4.10 e 4.11, caracterizaram os aspectos sobre as informações de GWS, os quais foram confrontados aos métodos RR-BLUP, Bayes A, Bayes B e BSSV com os métodos modificados, ou seja, RR-BLUP modificado, Bayes A modificado, Bayes B modificado e BSSV modificado. Neste caso, o valor absoluto do efeito de marcador do tipo SNP foi a medida apropriada, para a comparação dos efeitos estimados (eixo Y no gráfico *Manhattan plot*). As estimativas dos valores dos efeitos são apresentados nas Tabelas 2 a 25, em anexo.

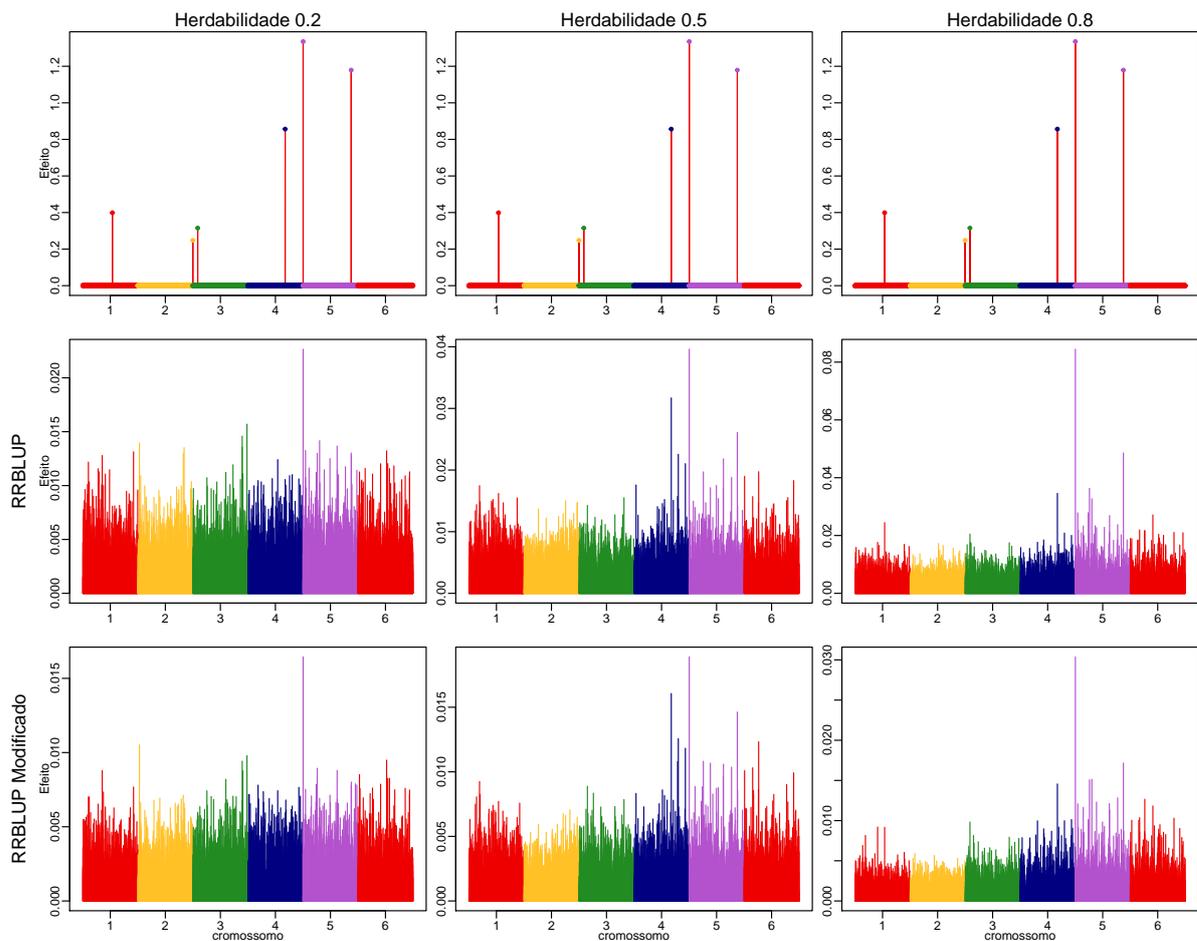
Para comparar os métodos modificados com a análise GWAS (implementada pelo pacote rrBLUP do *software R*), as Figuras 4.16, 4.17, 4.18, 4.19, 4.20, 4.21, 4.22 e 4.23 identificam os valores de efeito dos marcadores em associação com os genes simulados pelo gráfico *Manhattan plot*. O eixo Y representa o $-\log_{10}(\text{valor} - p)$ ou $\text{valor} - p$ do teste de Wald. O eixo X, representado pelo SNP, dentro de seu respectivo cromossomo. Em cada cromossomo, os SNPs são ordenados por marca e posição. Os valores do $-\log_{10}(\text{valor} - p)$ e $\text{valor} - p$ são apre-

sentados nas Tabelas 26 a 49, em anexo. Foram estabelecidas linhas de corte tracejadas, para identificar os SNPs selecionados, usando valores $-\log_{10}(0,05/k) = 5,0791$, com $k = 6000$ e $valor - p = 3,84$ para o teste de Wald.

Os métodos modificados também foram comparados aos métodos FASTmrMLM e FASTmrEMMA de análise multiloci (implementados pelo pacote mrMLM do *software R*), nas Figuras 4.24, 4.25, 4.26, 4.27, 4.28, 4.29, 4.30 e 4.31, identificados os valores de efeito dos marcadores em associação com os genes simulados, por meio do gráfico *Manhattan plot*. Foram estabelecidas linhas de corte tracejadas para identificar os SNPs selecionados, usando valores $-\log_{10}(0,05/k) = 5,0791$, com $k = 6000$ e $valor - p = 3,84$ para o teste de Wald.

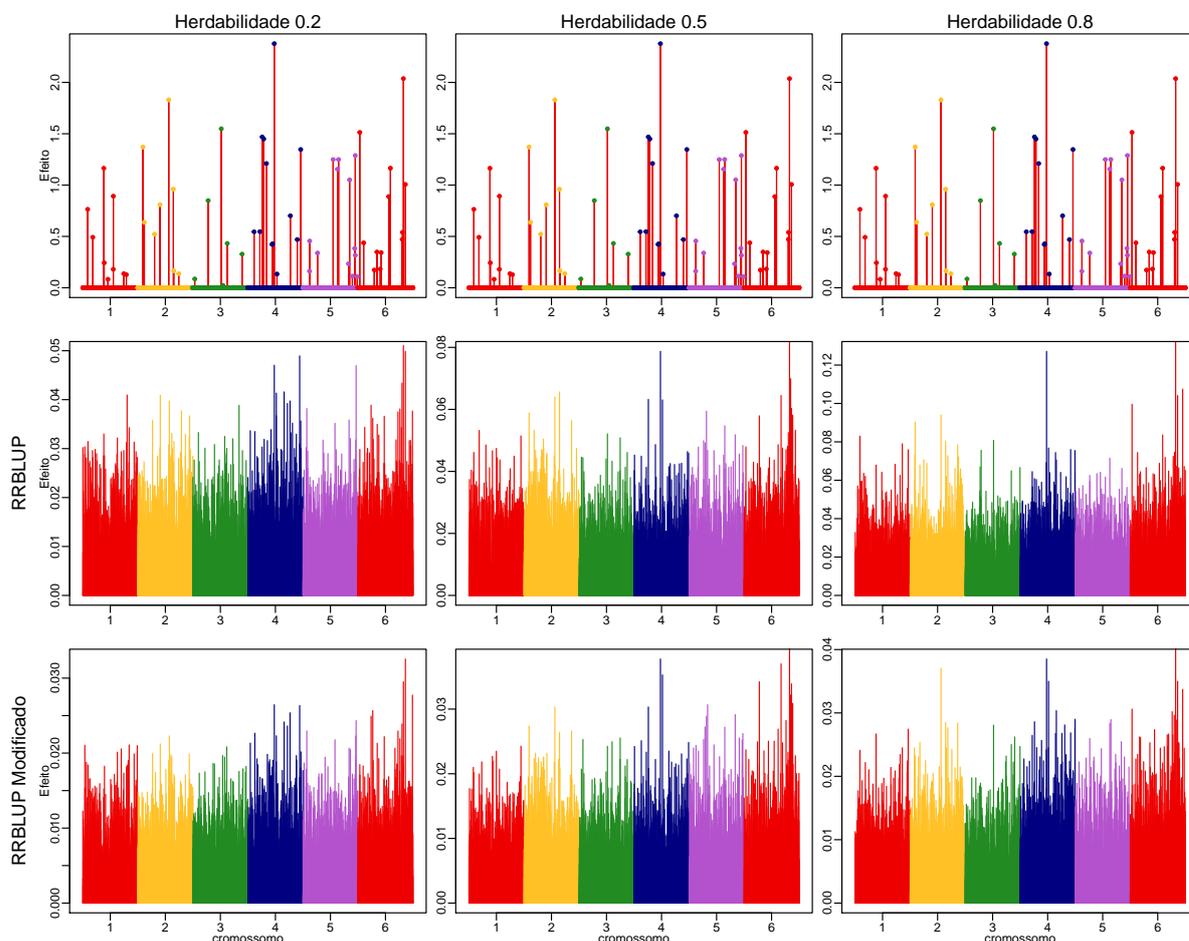
4.1.3.1 Análise Comparativa RR-BLUP versus RR-BLUP modificado

Figura 4.4 – Cenário Oligogênico - Painéis demonstrando a análise dos efeitos dos marcadores, para cada uma das herdabilidades $h^2 \in \{0,2;0,5;0,8\}$, confrontando o método RR-BLUP para GWS (pacote rrBLUP *software R*) e método RR-BLUP modificado



Fonte: Da autora (2019)

Figura 4.5 – Cenário Poligênico - Painéis demonstrando a análise dos efeitos dos marcadores, para cada uma das herdabilidades $h^2 \in \{0,2; 0,5; 0,8\}$, confrontando o método RR-BLUP para GWS (pacote *rrBLUP software R*) e método RR-BLUP modificado



Fonte: Da autora (2019)

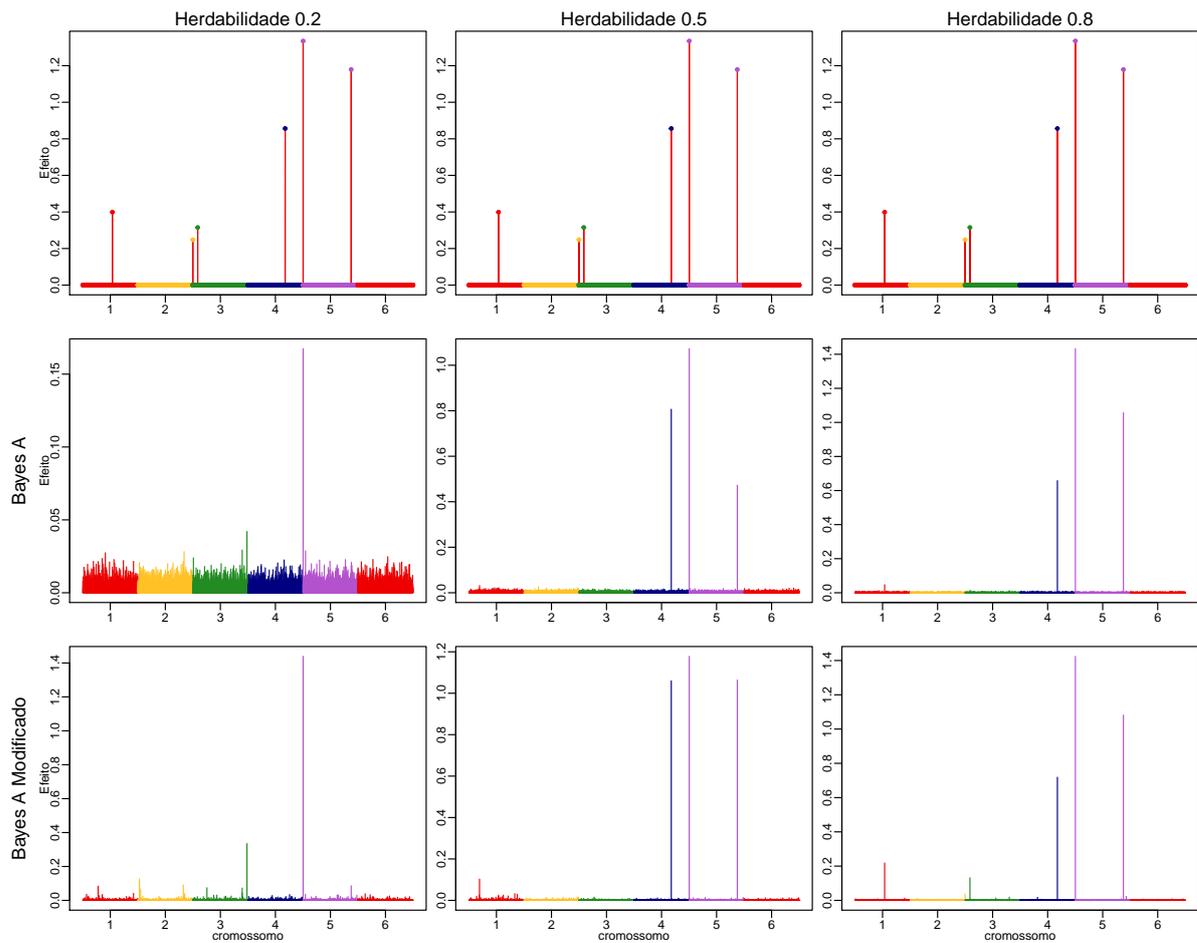
Como observado nas Figuras 4.4 e 4.5, os métodos RR-BLUP e RR-BLUP modificado apresentam equiparáveis estimativas dos efeitos dos marcadores SNPs. Para os dois métodos, em ambos cenários, as estimativas de efeito dos marcadores SNPs indicam-se muito próximas, que decorre do encolhimento do efeito do marcador. Esse encolhimento é inerente aos dois métodos, ou seja, para o método RR-BLUP modificado, o fator λ também é reconhecido como parâmetro de encolhimento e a ele atribuída a implantação de uma única variância para todas as marcas.

Pode-se afirmar que as duas abordagens representadas pelos dois métodos RR-BLUP e RR-BLUP modificado foram equivalentes e forneceram uma classificação muito próxima dos efeitos simulados para as análises (ver Figura 4.4). Essa equivalência depende de suposições que cercam a variância dos efeitos dos marcadores. Para implementação do RR-BLUP modificado, definiu-se a variância do marcador como $\frac{v_y}{k}$ (variância fenotípica sobre o número de

marcadores). Quanto à predição genômica, ambas as abordagens podem ser consideradas para análise.

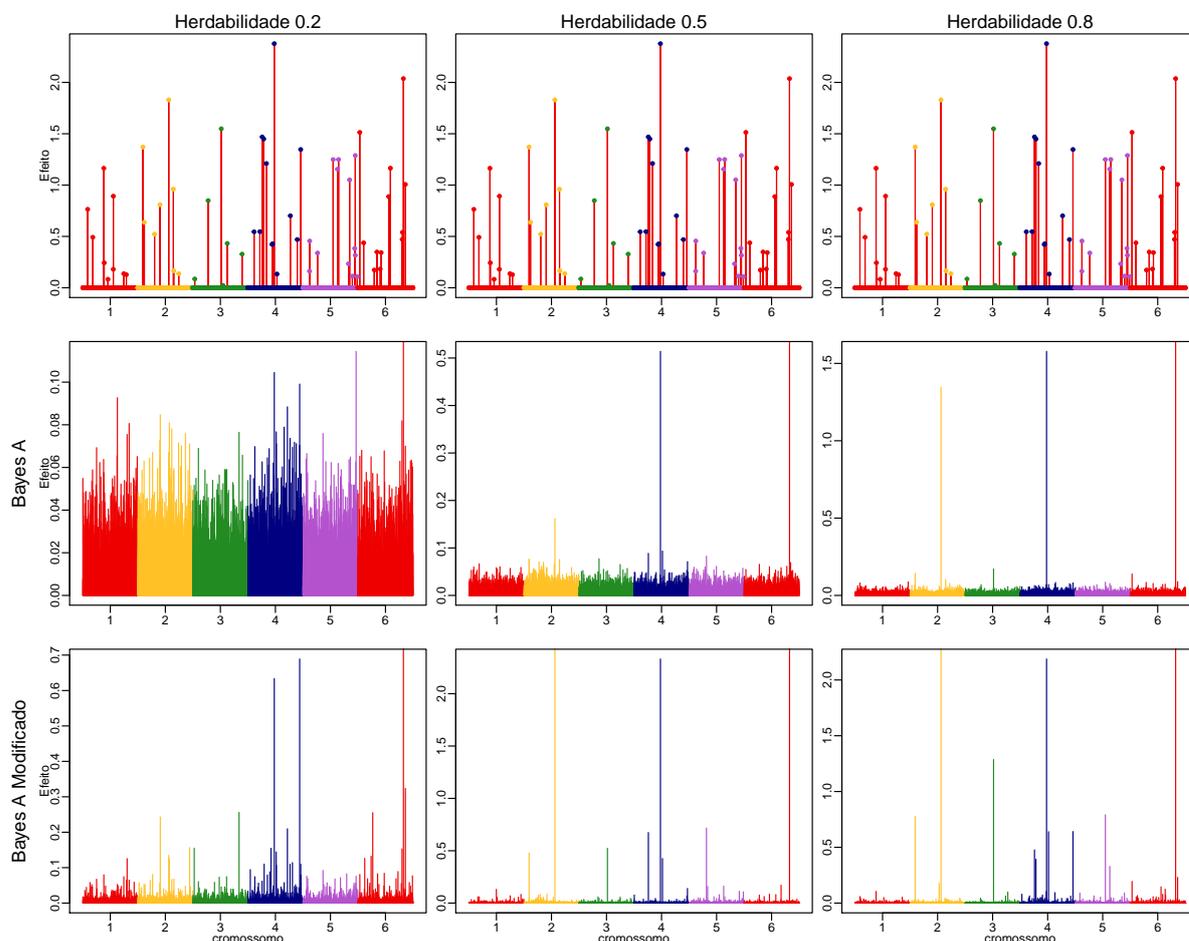
4.1.3.2 Análise Comparativa Bayes A versus Bayes A modificado

Figura 4.6 – Cenário Oligogênico - Painéis demonstrando a análise dos efeitos dos marcadores, para cada uma das herdabilidades $h^2 \in \{0,2;0,5;0,8\}$, confrontando o método Bayes A para GWS (pacote BGLR *software* R) com o método Bayes A modificado



Fonte: Da autora (2019)

Figura 4.7 – Cenário Poligênico - Painéis demonstrando a análise dos efeitos dos marcadores, para cada uma das herdabilidades $h^2 \in \{0,2; 0,5; 0,8\}$, confrontando o método Bayes A para GWS (pacote BGLR *software* R), com o método Bayes A modificado



Fonte: Da autora (2019)

Com relação ao método Bayes A, observa-se pelo *Manhattan plot* (Figuras 4.6 e 4.7) em comparação ao Bayes A modificado, que os efeitos das marcas apresentaram encolhimento mais próximo a zero, para todos os cenários, exceto cenário oligogênico com coeficientes de herdabilidades 0,5 e 0,8, no qual se equipararam em magnitude de escala. O encolhimento pode significar que um número menor de marcas tem efeito no fenótipo. Observa-se que, dentro de um mesmo cromossomo, os efeitos da marca pelo Bayes A modificado são observados como picos acentuados e mais espaçados entre si. Já pelo método Bayes A, por alto encolhimento, as marcas estão muito próximas, formando blocos em vez de torres como sugere o gráfico *Manhattan*, o que exibe uma possível inibição no potencial de detecção dos efeitos dos marcadores SNPs. Por outro lado, outras marcas pelo método Bayes A se sobressaem, ou seja, apresentaram alto efeito, porém, ao longo de toda a sequência, ficando mais espaçadas uma das outras.

Para a característica simulada com coeficiente de herdabilidade 0,8 e cenário oligogênico da Figura 4.6, identificou-se pelo método Bayes A modificado a presença de dois efeitos correspondentes ao gene de interesse e que não foram de magnitude expressiva para o método Bayes A. Uma ocorrência foi identificada no cromossomo dois (marca 2000, com efeito 0,03649) e outra no cromossomo três (marca 2088, com efeito 0,132238), além de vários outros de efeitos, evidentes em visualização gráfica pela magnitude dos efeitos (marcas 2983, 4047, 279, 922, ver Tabela 8). Como tal, observou-se indicação de um efeito correspondente ao gene de interesse, no final do cromossomo um (marca 538, com respectivo gene de interesse), assim como no cromossomo quatro (marca 3677) que apresenta um efeito de magnitude expressiva em concordância com a magnitude do gene de interesse simulado na parte inicial do cromossomo cinco (marca 4004) e na parte final do cromossomo cinco (marca 4877), também, ocorreu a presença de SNP que teve magnitude de correspondência com o efeito de gene simulado. Estes três últimos resultados de efeitos dos marcadores (marcas: 2088; 3677; 4004; 4877) foram os únicos equiparáveis aos resultados obtidos pelo método Bayes A e que fazem referência ao efeito do gene de interesse simulado (ver Tabela 10). Destaca-se que o efeito da marca 2088 (valor efeito 0,02531) pela análise do método Bayes A, não fica evidente pelo gráfico. Os resultados indicam que o método proposto, Bayes A modificado, é capaz de associar os efeitos com maior relevância aos fenótipos, para o cenário em questão.

Para os coeficientes de herdabilidade 0,2 e 0,5, no cenário oligogênico, os genes simulados tiveram menor quantidade de efeitos capturados correspondentes ao efeito estimado dos SNPs, quando comparados ao *Manhattan plot* com herdabilidade 0,80, ainda para o método Bayes A modificado. Porém, a eficiência na identificação do gene de interesse em associação significativa com o SNP estimado foi semelhante entre os fenótipos com diferentes herdabilidades para os dois métodos. O método Bayes A modificado, com herdabilidade 0,2, mostrou indício de ocorrência de um SNP de grande efeito, no terceiro cromossomo (marca 2983, com efeito 0,0376), não faz correspondência com o gene de efeito simulado, mas também apresentou magnitude expressiva pelo método Bayes A (com efeito 0,3347). Além disso, foi encontrado a presença de gene no cromossomo 3 (marca 4004), em ambos os métodos, pela manifestação da magnitude do efeito estimado para cada SNP. Na parte final do cromossomo cinco (marca 4877), que é um locus com presença de genes simulado, ocorreu o efeito de forma expressiva (no topo do *ranking* das marcas), considerando o efeito médio desta marca (0,086438), como o quinto valor de efeito obtido, dentre todos os efeitos estimados (Consultar Tabelas 8 e 9).

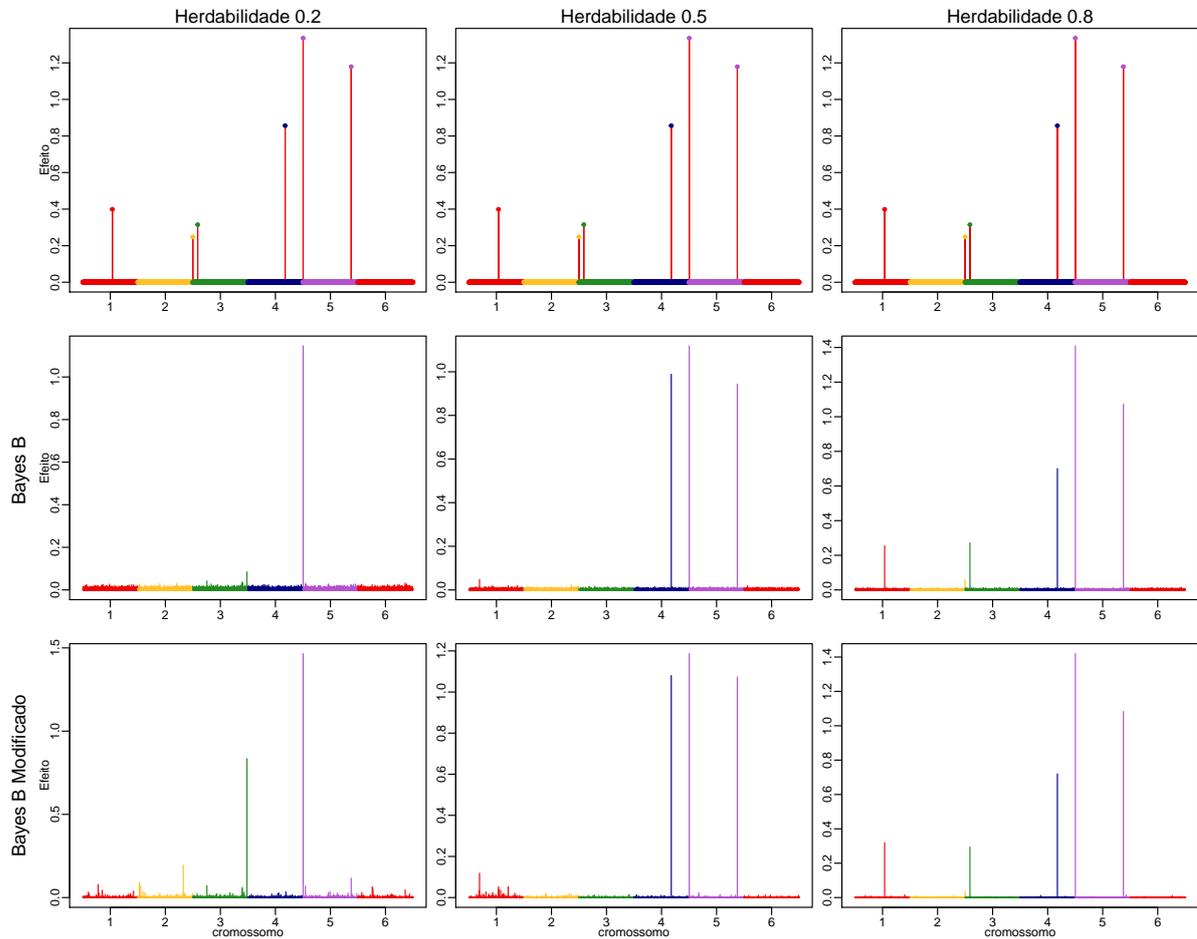
Outros indícios de ocorrência de SNPs com grande efeito, pelo método Bayes A modificado com herdabilidade 0,5 diz respeito às marcas 4004, 4877 e 3677 detectadas como associação ao locus de genes de interesse (Tabela 9). E, neste caso, a correspondência também foi equiparável ao método Bayes A. A marca 538 foi selecionada com efeito médio baixo (0,02581), porém permanece na ordem da oitava média capturada pelo ordenamento decrescente dos valores do efeito de cada marcador, pelo método Bayes A modificado e o mesmo não ocorreu como efeito de marca selecionado pelo método Bayes A.

Para o cenário poligênico, de coeficiente de herdabilidade 0,2, os gráficos de *Manhattan* (Figura 4.7), para efeito dos marcadores, apresentou um encolhimento evidentemente menor para efeito dos SNPs sob o método Bayes A em comparação com as condições do método Bayes A modificado. Deste fato, sete SNPs foram detectados (marca 382; 745;1562;3480; 5826; 5862), que representaram 11% da variação fenotípica. Na condição do método Bayes A, um total de três SNPs (marca 3941; 5826; 5862) apresentou magnitude expressiva, com relação aos genes de efeito simulado, o que explicou 5% da variação fenotípica. Na análise combinada, 60 genes foram simulados como representativos de características genéticas.

Nota-se, no entanto que, nas análises pelo Bayes A e Bayes A modificado com herdabilidades 0,5 e 0,8, também, apresentou-se um grande número de efeitos dos SNPs pelo gráfico *Manhattan plot* com amplitudes próximas de zero, fazendo correspondência a um pequeno número de genes em comparação aos 60 genes simulados.

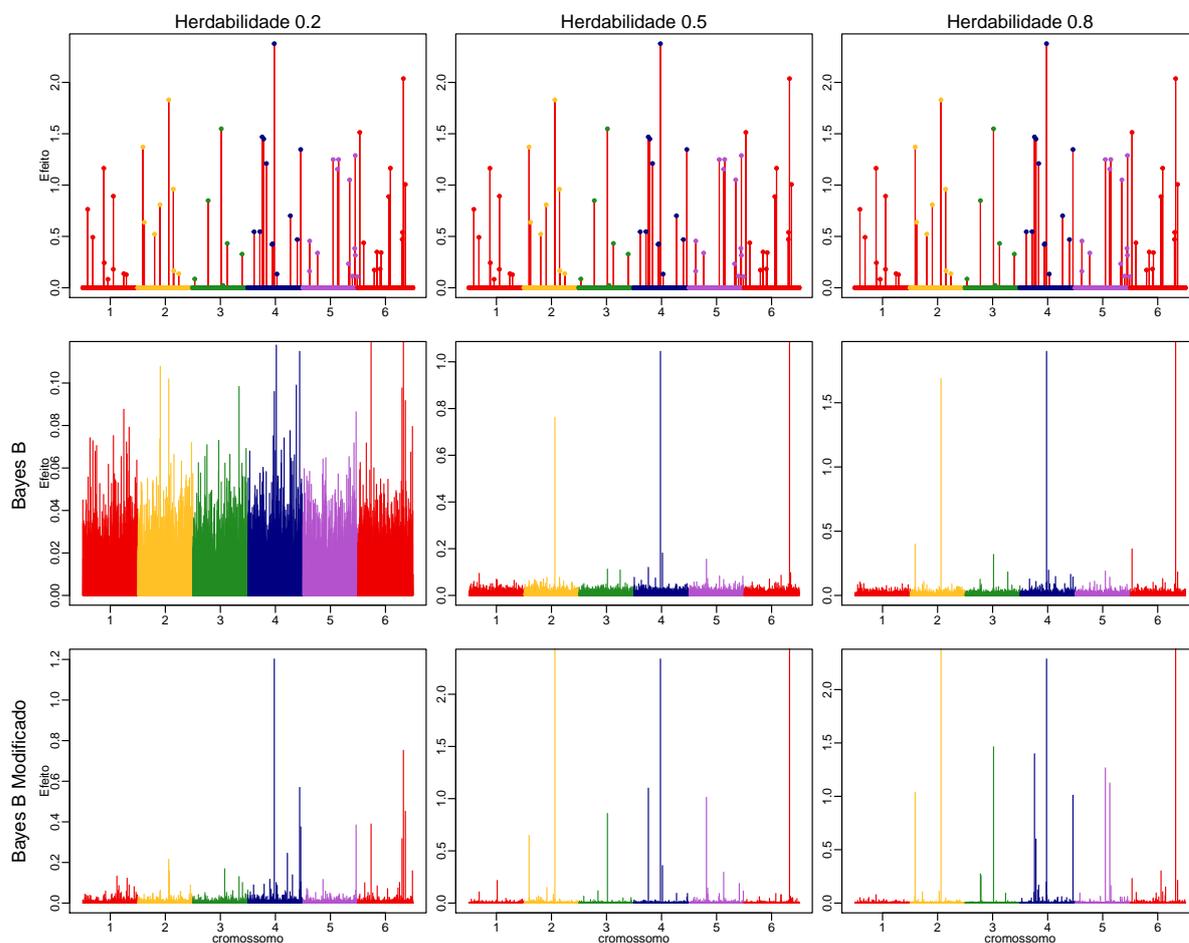
4.1.3.3 Análise Comparativa Bayes B versus Bayes B modificado

Figura 4.8 – Cenário Oligogênico - Painéis demonstrando a análise dos efeitos dos marcadores, para cada uma das herdabilidades $h^2 \in \{0,2;0,5;0,8\}$, confrontando o método Bayes B para GWS (pacote BGLR *software* R), com o método Bayes B modificado



Fonte: Da autora (2019)

Figura 4.9 – Cenário Poligênico - Painéis demonstrando a análise dos efeitos dos marcadores, para cada uma das herdabilidades $h^2 \in \{0,2; 0,5; 0,8\}$, confrontando o método Bayes B para GWS (pacote BGLR *software* R), com o método Bayes B modificado



Fonte: Da autora (2019)

Os resultados de efeitos dos marcadores são exibidos com os gráficos de *Manhattan plot* nas Figuras 4.8 e 4.9, em que se figuram os pontos dos SNPs, alguns com distribuição dos valores com magnitude maiores, comparando as análises pelo método Bayes B e Bayes B modificado. Um padrão interessante é que o Bayes B modificado gerou um número superior de marcadores que corresponderam aos genes simulados. Este aumento no número de marcadores com amplitude dos efeitos dos SNPs maiores também é percebido com o aumento no coeficiente de herdabilidade.

Uma inspeção na Figura 4.8 sobre o efeito dos marcadores estimados, mostrou que, para os métodos de Bayes B e Bayes B modificado, no cenário oligogênico 0,8, todos os marcadores contribuíram com o efeito expressivo aos seis genes simulados, observada a magnitude de encolhimento a zero maior pelo método Bayes B. Para os demais cenários, teve-se amplitudes expressivas, de acordo com as estimativas do efeito dos marcadores, para cada característica

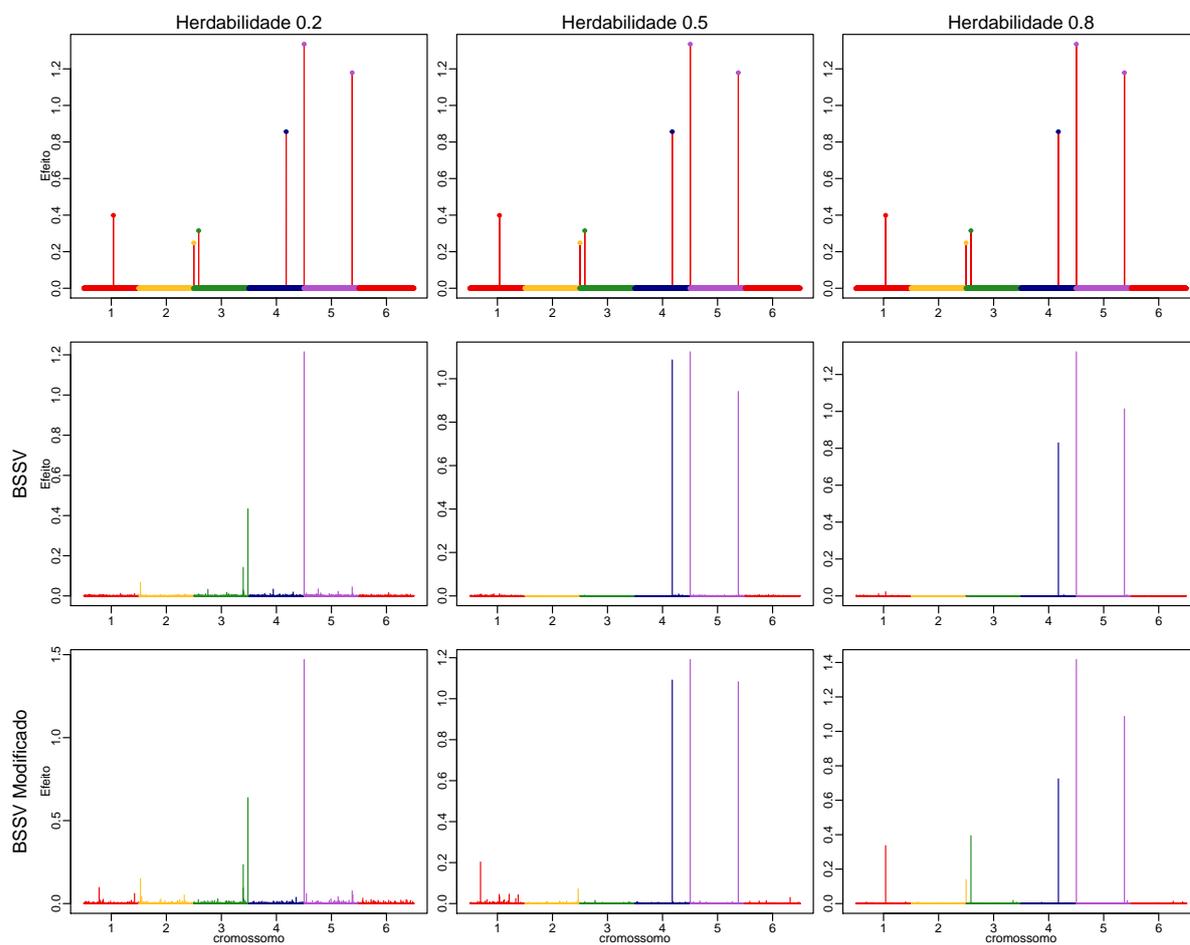
simulada, usando método Bayes B modificado com herdabilidade 0,2. As marcas foram 4877; 4004, localizadas no cromossomo cinco contra o método Bayes B, contra a única marca 4004. Para o coeficiente de herdabilidade 0,5, as marcas foram 538; 3677; 4004 e 4877, considerando o método Bayes B modificado, contra três marcas 3677; 4004 e 4877, pelo Bayes B.

Notou-se que nas análises de Bayes B e Bayes B modificado, com herdabilidades 0,5 e 0,8, também obteve-se expressiva magnitude de correspondência dos SNPs com os genes simulados (ver Figura 4.8). O número de genes expressivos foi maior para a análise, usando o método Bayes B modificado, em todos os cenários, exceto para coeficiente de herdabilidade 0,8, o qual manteve o total de seis efeitos.

Na observação referente à Figura 4.9, alcançou-se uma quantidade de genes expressivos em cada herdabilidade, no cenário poligênico. Embora o efeito de encolhimento seja evidente pelo método Bayes B, a flexibilidade na identificação de um número maior de marcadores SNPs é reconhecida pela exploração gráfica, para o método Bayes B modificado, em todos os coeficientes de herdabilidade.

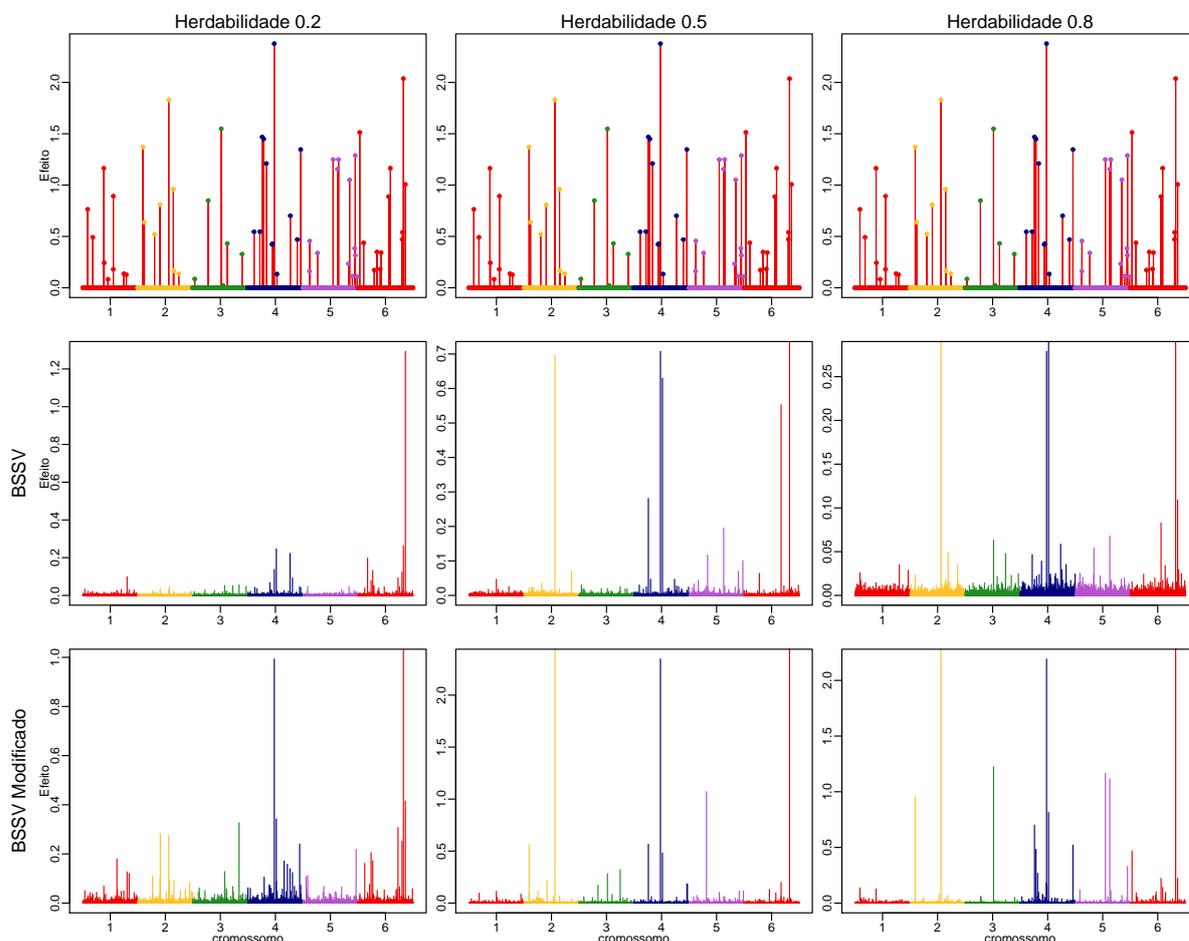
4.1.3.4 Análise Comparativa BSSV versus BSSV modificado

Figura 4.10 – Cenário Oligogênico - Painéis demonstrando a análise dos efeitos dos marcadores, para cada uma das herdabilidades $h^2 \in \{0,2;0,5;0,8\}$, confrontando o método BSSV, com o método BSSV modificado



Fonte: Da autora (2019)

Figura 4.11 – Cenário Poligênico- Painéis demonstrando a análise dos efeitos dos marcadores, para cada uma das herdabilidades $h^2 \in \{0,2; 0,5; 0,8\}$, confrontando o método BSSV, com o método BSSV modificado



Fonte: Da autora (2019)

Nas análises com o método *Bayesian Stochastic Search Variable BSSV* e pelo BSSV modificado, com coeficientes de herdabilidade 0,2 e 0,5, foram identificados os mesmos SNPs com valores de magnitude evidentes, comparados ao gene de interesse simulado (duas marcas 4004 e 4877 no grupo do cromossomo 5, com herdabilidade 0,2) e quatro marcas 4004; 4877; 3677; 538 nos cromossomos de 5, 4 e 1 para o coeficiente de herdabilidade 0,5 (Figura 4.10). Para análise com coeficiente de herdabilidade 0,8, foram reconhecidos todos os SNPs pelo método de BSSV modificado, contra cinco marcas de grande amplitude, pelo método BSSV (marcas 4004; 4877; 3677 e 538). Consultar Tabelas 20, 21 e 22, em anexo.

As estimativas dos efeitos de marcadores variaram, de acordo com a pressuposição de cada modelo genético, associado a cada metodologia de análise. Dentre os modelos bayesianos empregados, o BSSV é reconhecido pela grande capacidade de selecionar marcas de grande efeito em modelos com múltiplas marcas. Para os métodos propostos, tanto o método BSSV

quanto BSSV modificado mantiveram as características do modelo e apresentaram um comportamento de encolhimento muito próximo, a menos para o método BSSV, no cenário poligênico, nas herdabilidades 0,5 e 0,8. Para estes cenários, o encolhimento maior por parte do método BSSV é evidente e pode ser verificado pelas Figuras 4.10 e 4.11.

Os gráficos referentes ao cenário poligênico, com as devidas herdabilidade de 0,2, 0,5 e 0,8, são apresentados pela Figura 4.11. A análise realizada com o BSSV apresentou característica de encolhimento em magnitude das estimativas dos efeitos equiparáveis somente refere-se ao cenário de coeficiente de herdabilidade 0,2. Nos demais, retrata encolhimento maior comparado ao comportamento do método BSSV modificado. Assim, o BSSV modificado foi consistente em descrever os efeitos dos SNPs, no sentido de maior número de SNPS com efeitos em maior magnitude, compatíveis aos genes simulados.

Tabela 4.5 – Número de marcadores com valores de efeito relativamente moderados (ou maiores), detectados pelos diferentes métodos, dada a associação ao gene de interesse nos cenários oligogênico (6 genes simulados) e cenário poligênico (60 genes simulados)

| | | Número de marcadores SNPs | | | | | |
|---------|--------------------|---------------------------|---|---|------------|----|----|
| | | Oligogênico | | | Poligênico | | |
| métodos | RR-BLUP | 1 | 3 | 4 | 5 | 11 | 15 |
| | RR-BLUP modificado | 1 | 3 | 3 | 4 | 10 | 11 |
| | Bayes A | 1 | 3 | 5 | 3 | 16 | 18 |
| | Bayes A modificado | 2 | 4 | 6 | 7 | 12 | 21 |
| | Bayes B | 1 | 3 | 6 | 5 | 11 | 19 |
| | Bayes B modificado | 2 | 4 | 6 | 5 | 15 | 23 |
| | BSSV | 2 | 4 | 5 | 5 | 9 | 13 |
| | BSSV modificado | 2 | 4 | 6 | 6 | 14 | 23 |

Fonte: Da autora (2019)

Embora todos os métodos tenham padrões de contração diferentes, as consequências dos diferentes métodos tradicionais de seleção e seus perfis foram claramente refletidas e preservadas, em seus respectivos métodos modificados e, conseqüentemente, nos efeitos estimados dos SNPs. Como o comportamento da análise realizada com o BSSV e BSSV modificado, os quais apresentaram picos mais elevados, ao retratar os valores das estimativas dos marcadores, neste aspecto, implica respeitar a pressuposição dos métodos em controlar marcas de grande efeito. Ou ainda, o método RR-BLUP e RR-BLUP modificado apresentaram valores na escala muito próximos a zero, o que implica que todos os efeitos serão encolhidos em sua extensão. Ao mesmo tempo, fica clara a contribuição dos métodos modificados, com melhores resultados de identificação do gene, o que permite direcionar a real intenção de unificação com GWAS. Pode-

se ainda argumentar que o método BSSV modificado apresentou o melhor comportamento de seleção das variáveis ao direcionar fortemente o reconhecimento de efeito associado ao gene simulado. Depois do método BSSV modificado, a diferenciação de classificação fica Bayes B modificado, Bayes A modificado, sendo o RR-BLUP modificado e RR-BLUP métodos equiparáveis, porém os de menores eficiência. As diferenças, em relação aos métodos, podem ser observadas pelos resultados do número de marcadores detectados (Tabela 4.5).

Tabela 4.6 – Estimativas do coeficiente de herdabilidade \hat{h}^2 para cada um dos métodos nos cenários oligogênico e poligênico

| | | Estimativas da herdabilidade (\hat{h}^2) | | | | | |
|---------|--------------------|--|------|------|--------------------|------|------|
| | | cenário oligogênico | | | cenário poligênico | | |
| métodos | RR-BLUP | 0,23 | 0,52 | 0,80 | 0,22 | 0,52 | 0,82 |
| | RR-BLUP modificado | 0,18 | 0,39 | 0,62 | 0,14 | 0,35 | 0,51 |
| | Bayes A | 0,27 | 0,60 | 0,83 | 0,33 | 0,55 | 0,81 |
| | Bayes A modificado | 0,29 | 0,55 | 0,82 | 0,24 | 0,49 | 0,74 |
| | Bayes B | 0,28 | 0,52 | 0,85 | 0,28 | 0,57 | 0,80 |
| | Bayes B modificado | 0,29 | 0,55 | 0,82 | 0,23 | 0,50 | 0,71 |
| | BSSV | 0,20 | 0,49 | 0,77 | 0,13 | 0,35 | 0,52 |
| | BSSV modificado | 0,26 | 0,53 | 0,80 | 0,21 | 0,50 | 0,71 |
| | GWAS | 0,22 | 0,48 | 0,75 | 0,29 | 0,56 | 0,85 |

Fonte: Da autora (2019)

Na Tabela 4.6, as estimativas da herdabilidade, tanto nos cenários oligogênico quanto poligênico, foram, principalmente subestimadas pelo método RR-BLUP modificado e BSSV. As estimativas da herdabilidade, para os métodos RR-BLUP, Bayes A, Bayes A modificado, Bayes B e Bayes B modificado e BSSV, foram superestimadas para o cenário oligogênico, mas diminuíram, em alguns casos, no cenário poligênico, como para o método Bayes A modificado, para herdabilidades 0,49 e 0,74; Bayes B modificado com herdabilidade 0,71; BSSV com herdabilidade de 0,71. Para o método GWAS, há uma superestimativa para os dois cenários, exceto ao oligogênico com herdabilidade estimada em 0,48. Para o traço de baixa herdabilidade, o método Bayes A foi claramente superior no cenário poligênico, com herdabilidade 0,33. Uma outra observação é que o método BSSV produziu a menor estimativa da herdabilidade 0,13 e o terceiro menor $R^2 = 0,4882$ (ver Figura 4.15). Ocorreu melhor aproximação, nos valores de herdabilidade, para o método RR-BLUP (0,80), cenário oligogênico; Bayes B (0,80), cenário poligênico; Bayes B modificado (0,5) cenário poligênico.

Um fato verificado, para os dois cenários, em todos os métodos, é que o valor da PRESS decai com o aumento da herdabilidade, enquanto o R^2 aumenta com o aumento da herdabilidade

(Figuras 4.12, 4.13, 4.14 e 4.15). Observou-se que quando aumenta o coeficiente de herdabilidade, os valores de R^2 , em todos os métodos, aumentam para os dois cenários. Os dois primeiros métodos, RR-BLUP e RR-BLUP modificado, produziram estimativas do R^2 muito próximas, com diferenças, em alguns casos, a partir da segunda casa decimal, com valores menores para o modelo RR-BLUP modificado. Pelos valores da PRESS, acusaram-se menores valores para o método RR-BLUP em comparação com o RR-BLUP modificado, a menos no cenário oligogênico, herdabilidade 0,2. Juntos, R^2 e PRESS, permitem afirmar quanto a melhora do método RR-BLUP modificado, em comparação ao RR-BLUP, não ocorreu. Pela Figura 4.12, os gráficos ilustram a variação na resposta, para os dois métodos nas distintas herdabilidades, sugere variações com pontos mantendo o mesmo padrão, explicando métodos equiparáveis.

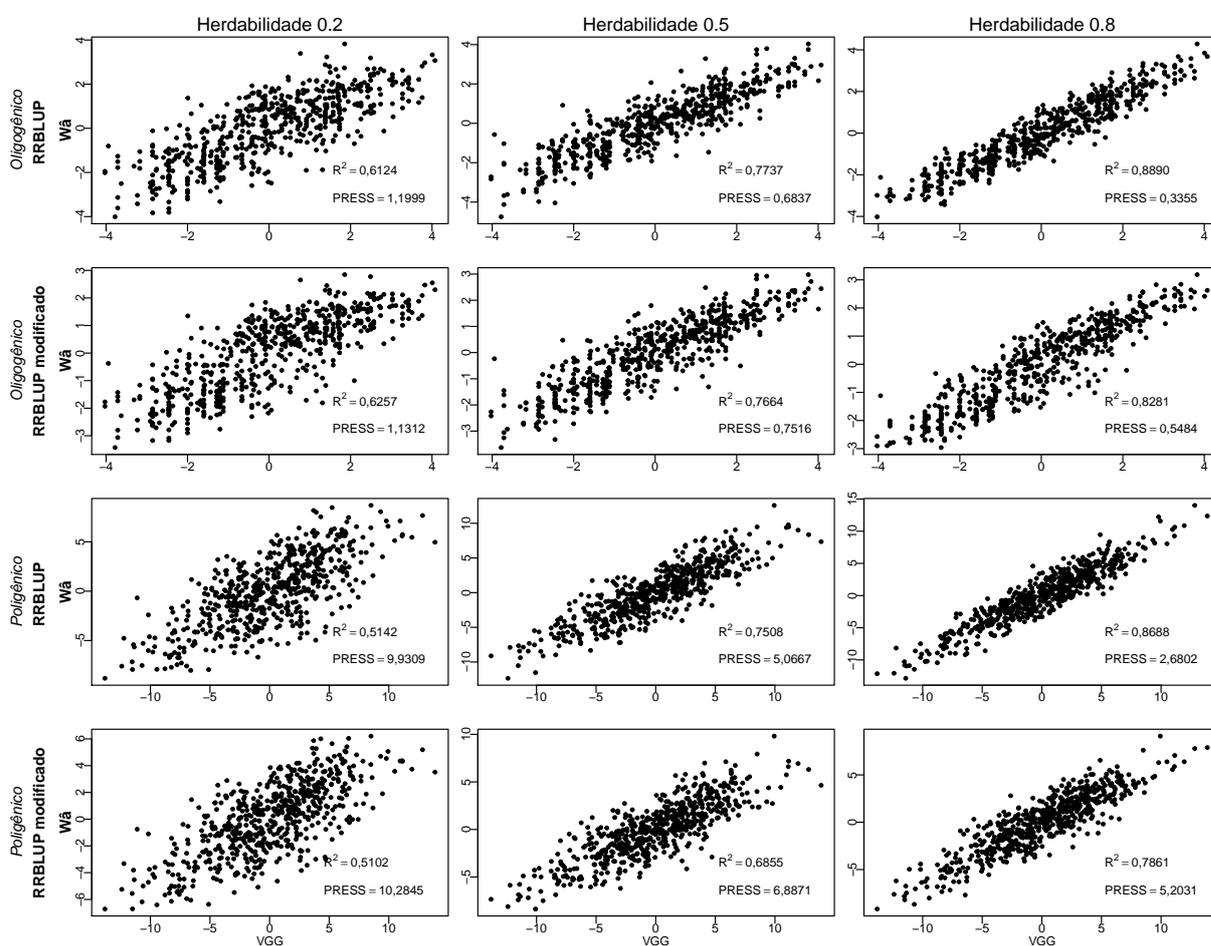
De forma geral, dentro de cada cenário, para os demais métodos, as estimativas de R^2 apresentaram-se maiores para os métodos modificados, Bayes A modificado e Bayes B modificado, determinando o quão bem os modelos predizem as respostas para as observações, conforme mostram as Figuras 4.13 e 4.14. O método BSSV modificado, em comparação ao BSSV, teve menores valores de R^2 (Figura 4.15).

Além disso, a estatística PRESS rendeu valores menores para os métodos Bayes A modificado e Bayes B modificado, para os dois cenários, em todos os coeficientes de herdabilidade, quando comparados aos respectivos métodos Bayes A e Bayes B. Sugere-se, portanto, que podem ocorrer melhoras consideráveis para estes métodos, se vierem a ser adotados como análises de seleção e associação.

Já o método RR-BLUP modificado apresentou maiores valores de PRESS (0,7516) para herdabilidade 0,2, PRESS (0,5484) para coeficiente de herdabilidade 0,5 no cenário oligogênico e também valores maiores da PRESS para o cenário poligênico, em todas as herdabilidades (ver Figura 4.12). O método BSSV modificado, fixada a herdabilidade 0,2 no cenário oligogênico, apresentou maior PRESS (0,8796) comparado ao método BSSV.

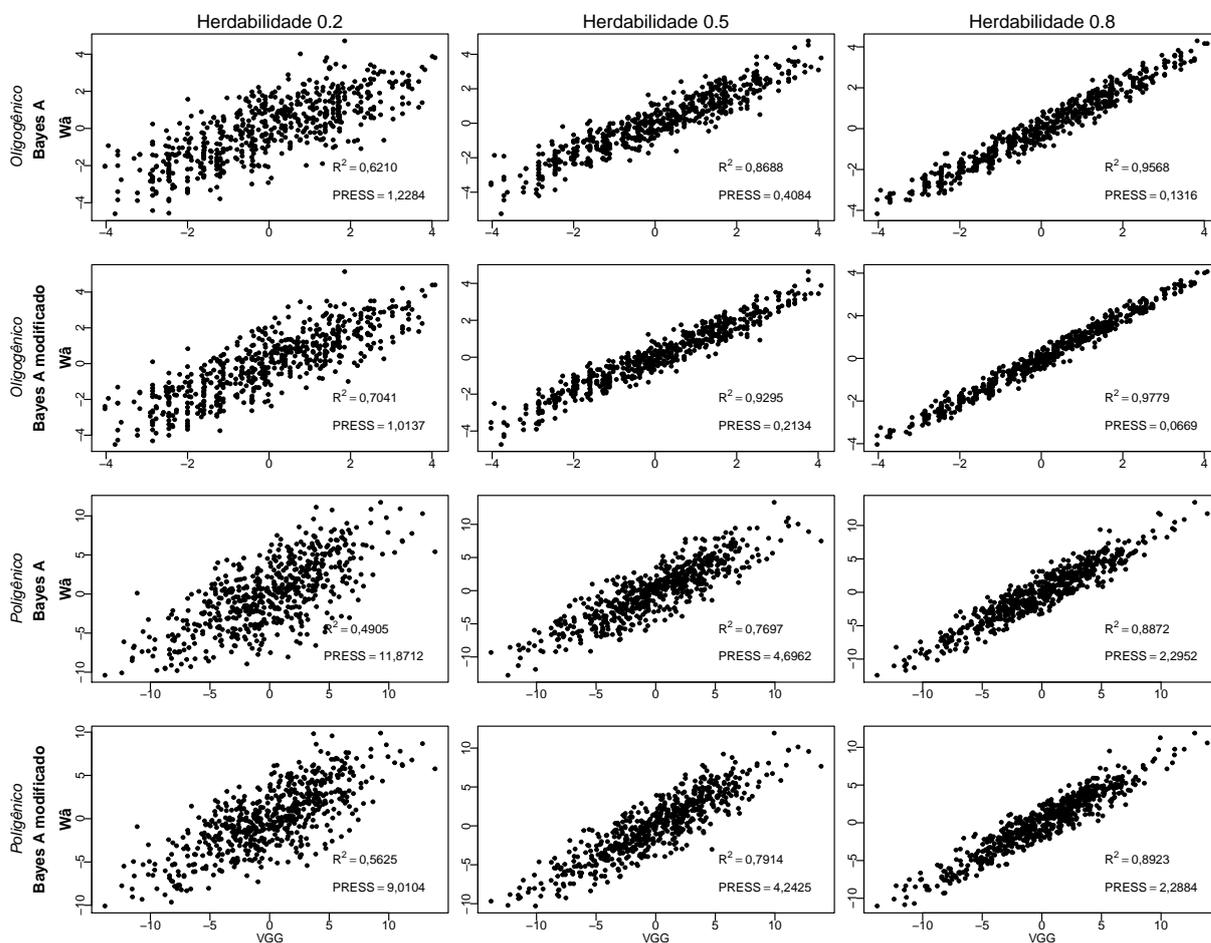
As Figuras 4.12, 4.13, 4.14 e 4.15 apresentam os gráficos de dispersão do valor genético genômico (VGG) verdadeiro contra valores genéticos estimados ($W\hat{a}$). Nos diagramas para cada herdabilidade, confrontando os métodos, nos cenários oligogênico e poligênico, observa-se a tendência de correlação positiva entre as variáveis.

Figura 4.12 – Painéis demonstrando a análise de dispersão entre o valor genético genômico (VGG) e valores genéticos estimados obtidos por $W\hat{a}$, para cada uma das herdabilidades $h^2 \in \{0,2; 0,5; 0,8\}$, confrontando o método RR-BLUP com o método RR-BLUP modificado, nos cenários oligogênico e poligênico. Estimativas do R^2 ajustado e PRESS entre VGG e $W\hat{a}$



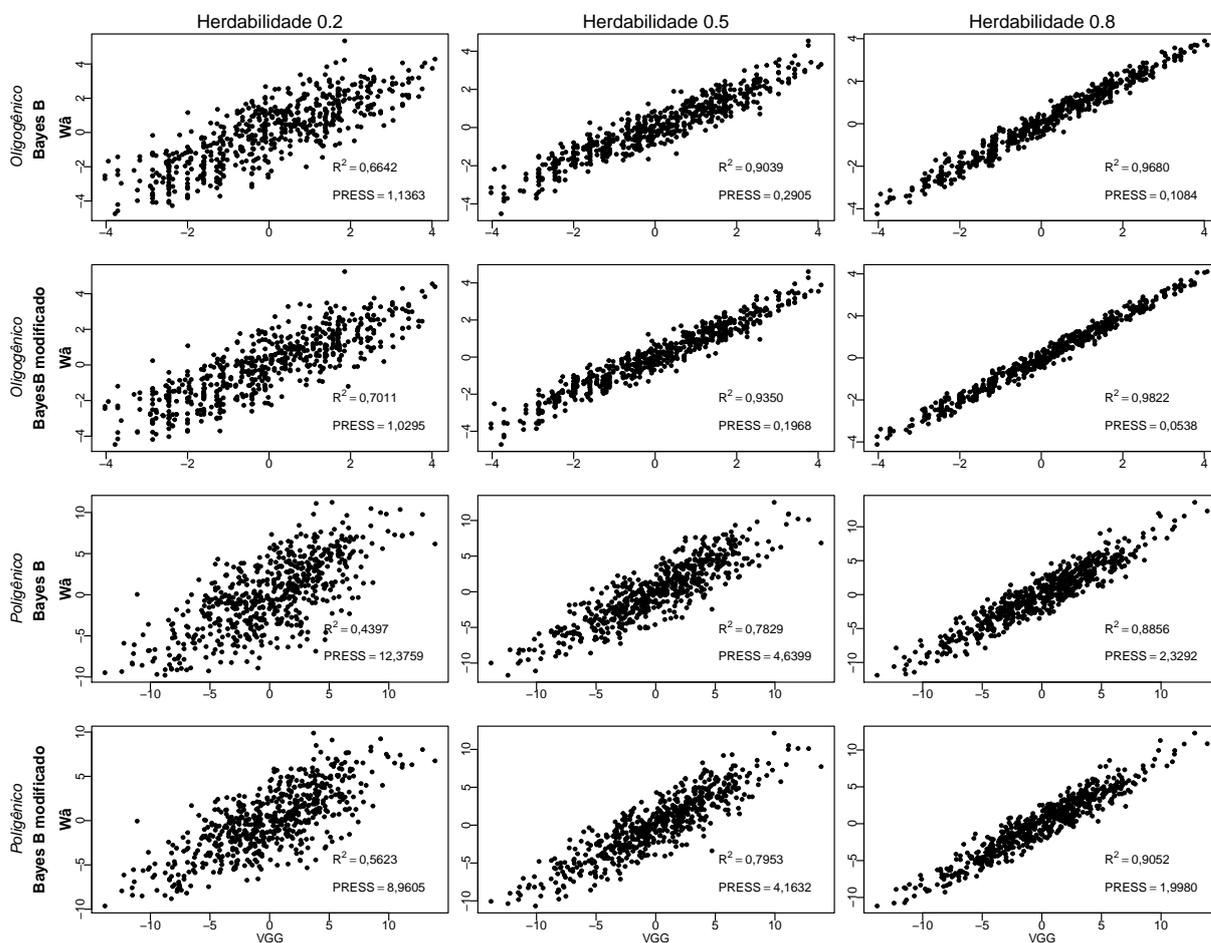
Fonte: Da autora (2019)

Figura 4.13 – Painéis demonstrando a análise de dispersão entre o valor genético genômico (VGG) e valores genéticos estimados obtidos por $W\hat{a}$, para cada uma das herdabilidades $h^2 \in \{0,2; 0,5; 0,8\}$, confrontando o método Bayes A com o método Bayes A modificado, nos cenários oligigênico e poligênico. Estimativas do R^2 ajustado e PRESS entre VGG e $W\hat{a}$



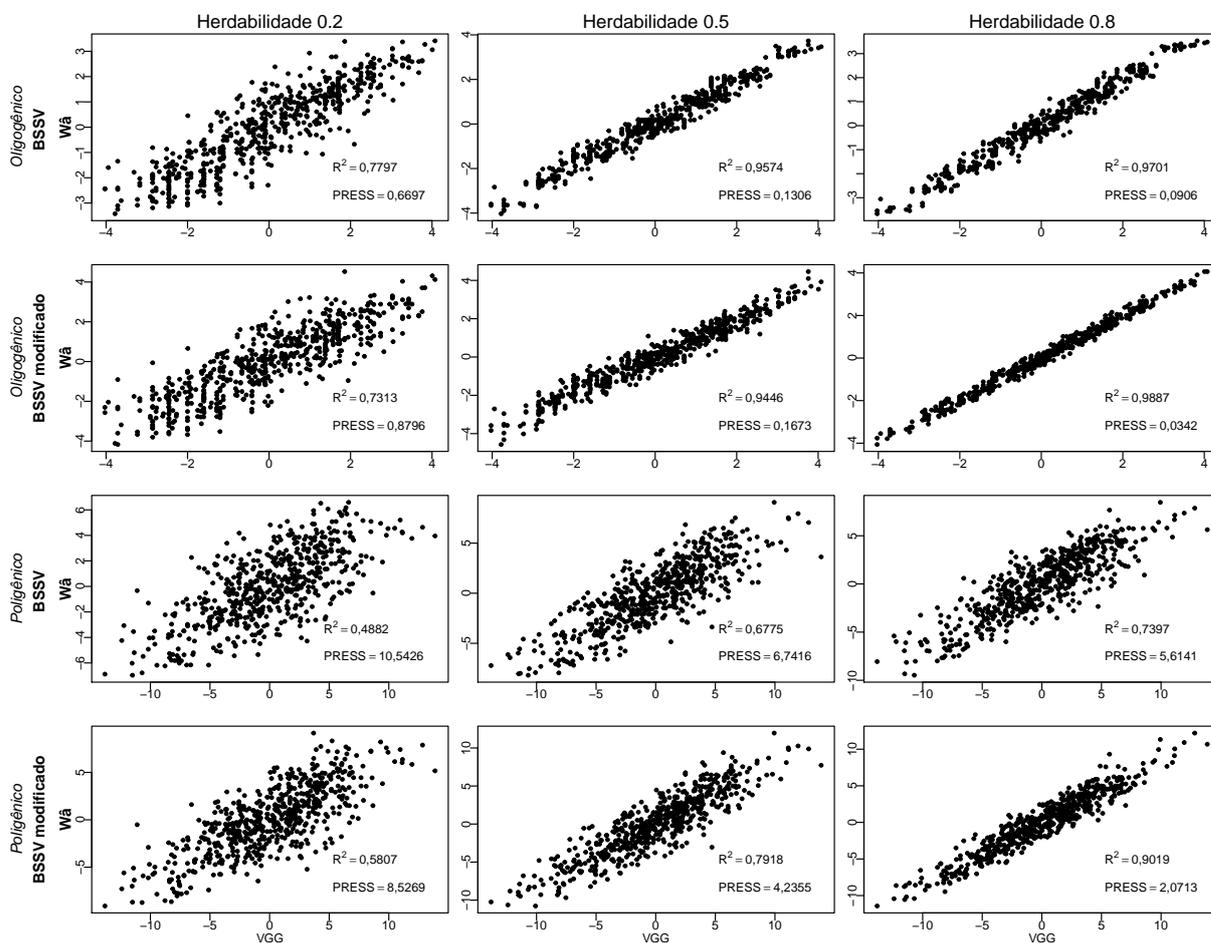
Fonte: Da autora (2019)

Figura 4.14 – Painéis demonstrando a análise de dispersão entre o valor genético genômico (VGG) e valores genéticos estimados obtidos por $W\hat{a}$, para cada uma das herdabilidades $h^2 \in \{0,2;0,5;0,8\}$, confrontando o método Bayes B com o método Bayes B modificado, nos cenários oligogênico e poligênico. Estimativas do R^2 ajustado e PRESS entre VGG e $W\hat{a}$



Fonte: Da autora (2019)

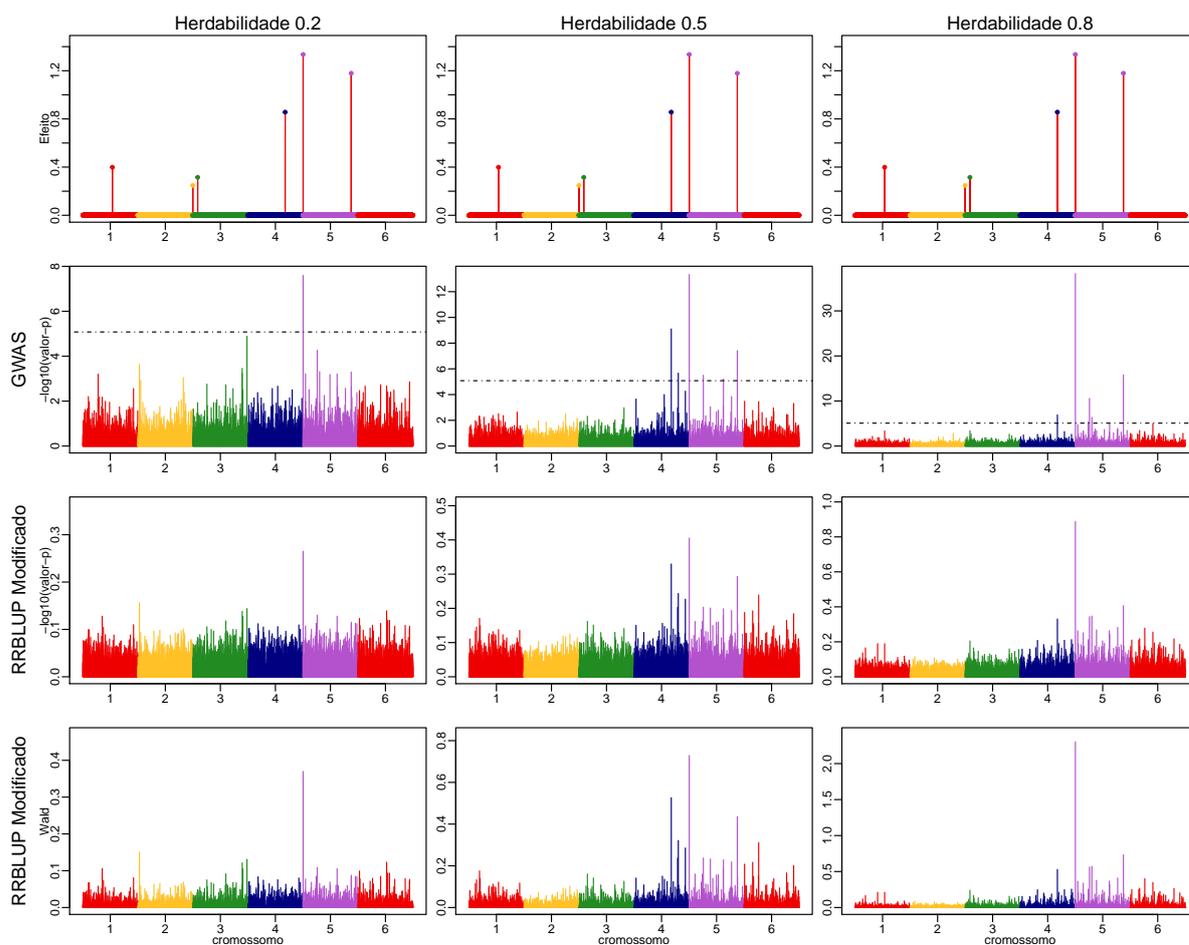
Figura 4.15 – Painéis demonstrando a análise de dispersão entre o valor genético genômico (VGG) e valores genéticos estimados obtidos por $W\hat{a}$, para cada uma das herdabilidades $h^2 \in \{0,2; 0,5; 0,8\}$, confrontando o método BSSV com o método BSSV modificado, nos cenários oligogênico e poligênico. Estimativas do R^2 ajustado e PRESS entre VGG e $W\hat{a}$



Fonte: Da autora (2019)

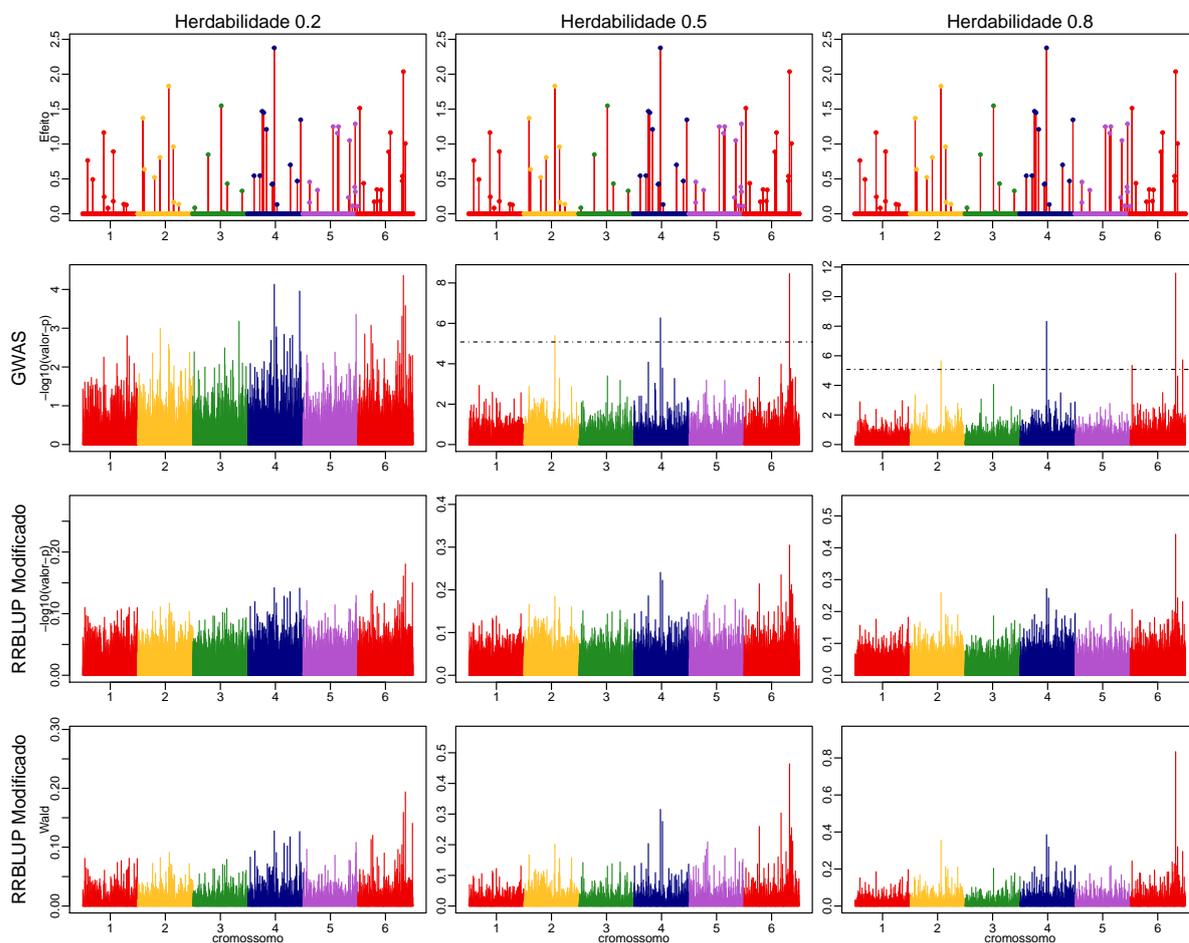
4.1.3.5 Análise Comparativa RR-BLUP modificado versus GWAS

Figura 4.16 – Cenário Oligogênico - Painéis demonstrando a análise dos efeitos dos marcadores para cada uma das herdabilidades $h^2 \in \{0,2;0,5;0,8\}$, confrontando o método GWAS (Manhattan plot de $-\log_{10}(\text{valor} - p)$), com o método RR-BLUP modificado (Manhattan plot de $-\log_{10}(\text{valor} - p)$ e $\text{valor} - p$ do teste Wald), para associação de 6000 SNPs com total de 6 cromossomos



Fonte: Da autora (2019)

Figura 4.17 – Cenário Poligênico - Painéis demonstrando a análise dos efeitos dos marcadores para cada uma das herdabilidades $h^2 \in \{0,2;0,5;0,8\}$, confrontando o método GWAS (Manhattan plot de $-\log_{10}(\text{valor} - p)$), com o método RR-BLUP modificado (Manhattan plot de $-\log_{10}(\text{valor} - p)$ e $\text{valor} - p$ de teste Wald), para associação de 6000 SNPs com total de 6 cromossomos



Fonte: Da autora (2019)

Tabela 4.7 – Detecção das marcas significativas por cromossomo. Resultados de associação genômica GWAS pacote rrBLUP

| GWAS | | | | | | | | | | |
|-------------|---------------|-------|-----------|--------------------------------|------------|---------------|-------|-----------|--------------------------------|-----------|
| cenário | Herdabilidade | Marca | Cr | $-\log_{10}(\text{valor} - p)$ | cenário | Herdabilidade | Marca | Cr | $-\log_{10}(\text{valor} - p)$ | |
| Oligogênico | 0,2 | 4004 | 5 | 7,597231* | Poligênico | 0,2 | - | - | - | |
| | | 4004 | 5 | 13,339721* | | | 0,5 | 5826 | 6 | 8,459814* |
| | | 3677 | 4 | 9,102989* | | | | 3480 | 4 | 6,266500* |
| | 4877 | 5 | 7,418931* | 1562 | | 2 | | 5,390297* | | |
| | 0,5 | 3805 | 4 | 5,668896** | | 0,8 | 5826 | 6 | 11,58792* | |
| | | 4260 | 5 | 5,507352** | | | 3480 | 4 | 8,31891* | |
| | | 4625 | 5 | 5,215285** | | | 5954 | 6 | 5,72248** | |
| | 0,8 | 4004 | 5 | 38,278105* | | 0,8 | 1562 | 2 | 5,66249* | |
| | | 4877 | 5 | 15,782135* | | | 5033 | 6 | 5,35015* | |
| | | 4260 | 5 | 10,589633** | | | | | | |
| | | | 3677 | 4 | | 6,920932* | | | | |
| | | | 4303 | 5 | | 6,386065** | | | | |
| | | 4237 | 5 | 5,372977** | | | | | | |

* valor verdadeiro positivo; ** falso positivo; Cr=cromossomo

Fonte: Da autora (2019)

As estimativas dos efeitos dos marcadores para características em que o método RR-BLUP modificado foi analisado, mostrou-se aparentemente igual ao método GWAS (Figuras 4.16 e 4.17). Considerou-se para o método RR-BLUP modificado, a adoção de uma única variância, para todas as marcas, a fim de realizar o teste de Wald (usando valores de corte $valor - p = 3,84$).

Para os dois métodos, em ambos os cenários, as estimativas de efeito dos marcadores SNPs aparentam uma disposição gráfica, a menos da magnitude, equiparável. Em todo caso, quando se direciona a observação ao método RR-BLUP modificado, verifica-se que nenhum SNP ultrapassou a limiar para associação, considerado a linha de corte do teste de Wald ($valor - p = 3,84$).

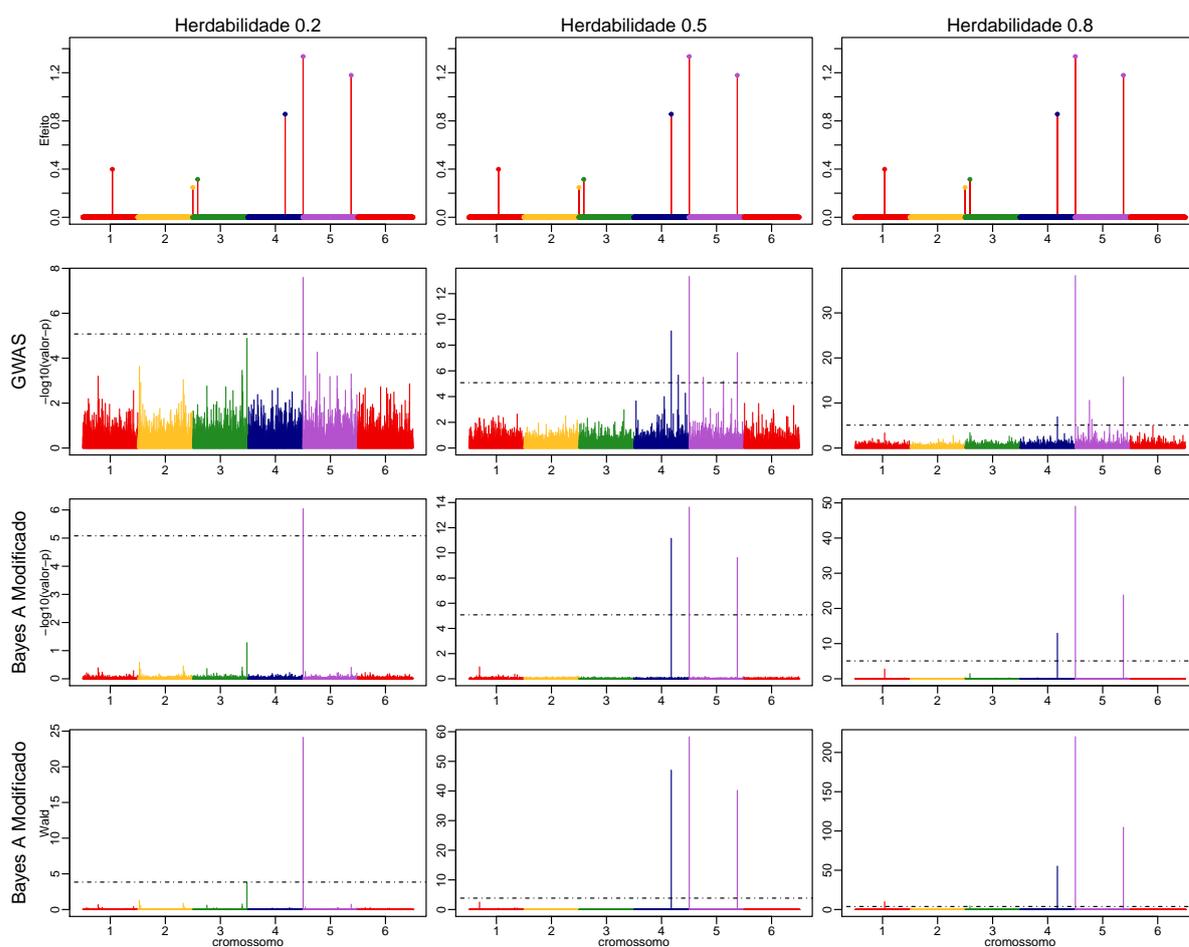
Quando se utilizou o método GWAS, para o estudo de associação genômica, analisando o cenário oligogênico (Figura 4.16), foi possível a identificação de uma marca significativa (estabelecida linha de corte de $-\log_{10}(0.05/k)$, localizada no cromossomo 5, marca 4004. O valor do $-\log_{10}(valor - p)$ dessa marca foi de, aproximadamente, 7,597231 para herdabilidade 0,2. Já para o método GWAS herdabilidade 0,5, observou-se a identificação de seis marcas significativas, das quais três apresentaram valores falso positivo, localizados nos cromossomos 4 e 5, marcas 3805, 4260 e 4625, cujos valores de $-\log_{10}(valor - p)$ foram 5,668896, 5,507352 e 5,215285, respectivamente (ver Tabela 4.7). Ainda na Figura 4.16, para o método GWAS herdabilidade 0,8, têm-se efeitos significativos para oito SNPs, porém cinco destes, apresentaram resultados de efeitos falsos positivos. Isto ocorre pela conseqüente correlação entre os SNPs. Percebe-se, contudo que o SNP mais significativo, considerando os três coeficientes de herdabilidade, foi o SNP localizado no cromossomo 5, marca 4004.

Para o cenário poligênico (Figura 4.17), pelo método GWAS e herdabilidade 0,5, somente três SNPs significativos foram localizados em três cromossomos (marcas 5826; 3489; 1562, em ordem de significância). Da mesma forma, para o método GWAS, considerando herdabilidade 0,8, somente quatro SNPs significativos com um SNP falso positivo, localizado no cromossomo 6, posição 5954, e $-\log_{10}(valor - p) = 5,722483$.

Comparando-se o poder de detecção de efeito significativo entre RR-BLUP modificado e GWAS, percebe-se que, mesmo apresentando falsos positivos, a GWAS mostrou um desempenho maior. Este poder é aumentado à medida que aumenta a herdabilidade.

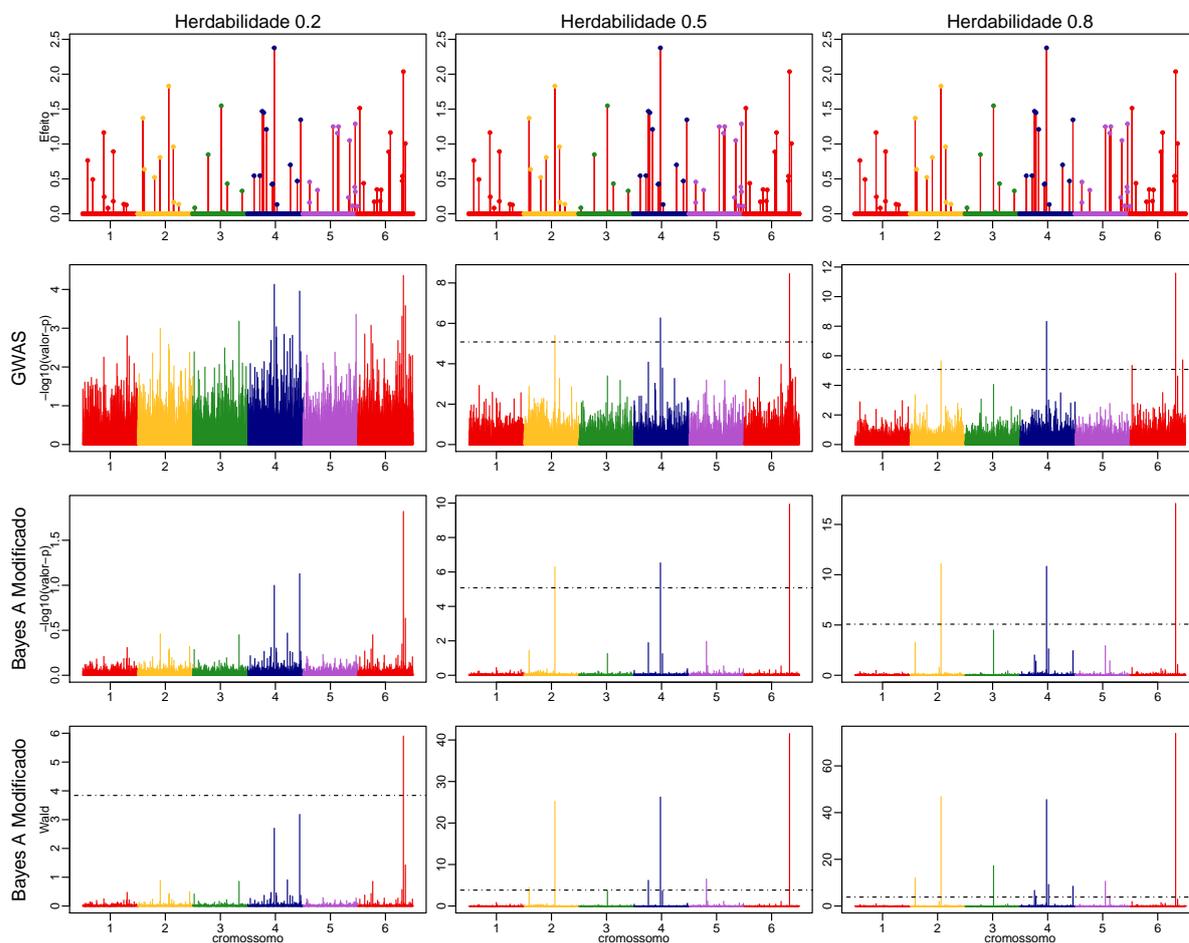
4.1.3.6 Análise Comparativa Bayes A modificado versus GWAS

Figura 4.18 – Cenário Oligogênico - Painéis demonstrando a análise dos efeitos dos marcadores, para cada uma das herdabilidades $h^2 \in \{0,2;0,5;0,8\}$, confrontando o método GWAS (Manhattan plot de $-\log_{10}(\text{valor} - p)$), com o método Bayes A modificado (Manhattan plot de $-\log_{10}(\text{valor} - p)$ e $\text{valor} - p$ de teste Wald), para associação de 6000 SNPs com total de 6 cromossomos



Fonte: Da autora (2019)

Figura 4.19 – Cenário Poligênico - Painéis demonstrando a análise dos efeitos dos marcadores, para cada uma das herdabilidades $h^2 \in \{0,2;0,5;0,8\}$, confrontando o método GWAS (Manhattan plot de $-\log_{10}(\text{valor} - p)$), com o método Bayes A modificado (Manhattan plot de $-\log_{10}(\text{valor} - p)$ e $\text{valor} - p$ de teste Wald), para associação de 6000 SNPs com total de 6 cromossomos



Fonte: Da autora (2019)

Tabela 4.8 – Detecção das marcas significativas por cromossomo. Resultados de associação genômica GWAS pacote rrBLUP e Método Bayes A modificado

| GWAS | | | | | cenário Poligênico | | | | | | |
|---------------------|-------|------|--------------------------------|--------------------------------|--------------------|-------------|------|--------------------------------|--------------------------------|-----------|--|
| cenário Oligogênico | | | | | cenário Poligênico | | | | | | |
| h^2 | Marca | Cr | $-\log_{10}(\text{valor} - p)$ | | h^2 | Marca | Cr | $-\log_{10}(\text{valor} - p)$ | | | |
| 0,2 | 4004 | 5 | 7,597231* | | 0,2 | - | - | - | | | |
| | 4004 | 5 | 13,339721* | | | 0,5 | 5826 | 6 | 8,459814* | | |
| | 3677 | 4 | 9,102989* | | | | 3480 | 4 | 6,266500* | | |
| | 4877 | 5 | 7,418931* | | | | 1562 | 2 | 5,390297* | | |
| | 0,5 | 3805 | 4 | 5,668896** | | 0,8 | 5826 | 6 | 11,58792* | | |
| | | 4260 | 5 | 5,507352** | | | 3480 | 4 | 8,31891* | | |
| 4625 | | 5 | 5,215285** | | 5954 | | 6 | 5,72248** | | | |
| 4004 | | 5 | 38,278105* | | 1562 | | 2 | 5,66249* | | | |
| 0,8 | 4877 | 5 | 15,782135* | | 0,8 | 5033 | 6 | 5,35015* | | | |
| | 4260 | 5 | 10,589633** | | | | | | | | |
| | 3677 | 4 | 6,920932* | | | | | | | | |
| | 4303 | 5 | 6,386065** | | | | | | | | |
| | 4237 | 5 | 5,372977** | | | | | | | | |
| Bayes A modificado | | | | | Cenário Poligênico | | | | | | |
| cenário Oligogênico | | | | | Cenário Poligênico | | | | | | |
| h^2 | Marca | Cr | Wald | $-\log_{10}(\text{valor} - p)$ | h^2 | Marca | Cr | Wald | $-\log_{10}(\text{valor} - p)$ | | |
| 0,2 | 4004 | 5 | 24,12597397* | 6,0446252* | 0,2 | 5826 | 6 | 5,903612* | 1,8207394 | | |
| | 4004 | 5 | 58,235291* | 13,633448* | | | 5826 | 6 | 41,549886* | 9,939676* | |
| 0,5 | 3677 | 4 | 47,019289* | 11,153097* | 0,5 | 3480 | 4 | 26,263244* | 6,525923* | | |
| | 4877 | 5 | 40,187383* | 9,636889* | | 1562 | 2 | 25,275985* | 6,303769* | | |
| 0,8 | 4004 | 5 | 219,999602* | 49,043530* | 0,8 | 4317 | 5 | 6,499547* | | | |
| | 4877 | 5 | 104,615975* | 23,828986* | | 3261 | 4 | 6,193938* | | | |
| | 3677 | 4 | 55,151248* | 12,952325* | | 1093 | 2 | 4,423353* | | | |
| | 538 | 1 | 9,892960* | | | 5826 | 6 | 73,9293430* | 17,091709* | | |
| | 2088 | 3 | 4,497320* | | | 1562 | 2 | 46,8210410* | 11,109166* | | |
| | | | | | | 3480 | 4 | 45,5758688* | 10,833157* | | |
| | | | | 2517 | 3 | 17,3014008* | | | | | |
| | | | | 1093 | 2 | 12,0016842 | | | | | |
| | | | | 4549 | 5 | 10,6081518 | | | | | |
| | | | | 3517 | 4 | 9,2493658** | | | | | |
| | | | | 3960 | 4 | 8,4597624* | | | | | |
| | | | | 3261 | 4 | 6,6959659* | | | | | |
| | | | | 4629 | 5 | 4,4155701* | | | | | |
| | | | | 3288 | 4 | 4,2546090* | | | | | |

* valor verdadeiro positivo; ** falso positivo; Cr=cromossomo

Fonte: Da autora (2019)

Comparando os resultados da análise bayesiana pelo método Bayes A modificado e GWAS, entre os dados dos genes simulados e valores-p dos efeitos marcadores estimados, para os cenários oligogênico e, para o coeficiente de herdabilidade 0,2, um SNP de marca 4004 foi comum a ambos os métodos e o único que alcançou o nível de significância para a associação genômica nas duas análises, com relação às três métricas $-\log_{10}(\text{valor} - p)$ e $\text{valor} - p$ do teste de Wald (Tabela 4.8). No cenário oligogênico, herdabilidade 0,5 observado na Figura 4.18, três SNPS (marcas 4004; 3677 e 4877) significativos foram comuns nas duas análises (Tabela 4.8).

Para herdabilidade 0,5 em GWAS, alguns SNPs tiveram um alto $-\log_{10}(\text{valor} - p)$, na região entre o cromossomo 4 e 5, vizinhos aos SNPs de associação significativa com o gene simulado (marcas 4004; 3677 e 4877), criando pequenas torres como estruturas do gráfico *Manhattan* e funcionando como suporte aos SNPs significativos (como se vê na Figura 4.18). Nestes picos, sugere-se, em muitos casos, procurar por pontos de associação genômica significativa,

mas, por outro lado, muitos destes SNPs significativos são falsos positivos, como os três SNPs de marcas 3805; 4260; 4625.

Para o coeficiente de herdabilidade 0,8, no cenário oligogênico, as análises pelo método Bayes A modificado apresentaram algo apropriado, com cinco SNPs significativos para associação genômica (marcas 4004; 4877; 3677; 538 e 2088), no total de seis efeitos de genes simulados. Comparando ao GWAS, os três SNPs significativos, localizados entre cromossomo 4, 5 e 6, em um total de oito SNPs significativos, cinco foram falsos positivos (marcas 4260; 4303; 4237; 5412; 4625). Observou-se, no entanto, que, mesmo identificando as regiões com marcas de associação genômica, ainda assim, a taxa de falsos positivos pelo método Bayes A modificado foi baixa (zero falsos positivos). De fato, a abordagem estatística pelo método Bayes A modificado manifestou maior sentido aos dados analisados, podendo controlar associações falso positivas e evitar associações falso negativas (no caso, uma única marca 2000 acusando falso negativo para Bayes A modificado, contra três pelo método GWAS).

Como esperado, o número de SNPs significativos foi melhor captado quando corrigimos a estrutura populacional e de parentesco, usando o método Bayes A modificado. Os gráficos da Figura 4.18 mostram uma melhoria clara pelo método Bayes A modificado.

Quando combinada a análise para o cenário poligênico, Figura 4.19, os melhores resultados em associação foram obtidos pela análise do método Bayes A modificado. O que ocorreu foi uma associação significativa, para a marca 5826, com coeficiente de herdabilidade 0,2, contra zero associações significativas para GWAS. Já para o coeficiente de herdabilidade 0,5, verificaram seis associações significativas para as marcas 5826; 3480; 1562; 4317; 3261 e 1093 com *valores - p* pelo teste de Wald, dos quais as três primeiras marcas resultaram em associação genômica significativa com $-\log_{10}(\text{valor} - p)$ de Wald, pelo método Bayes A modificado.

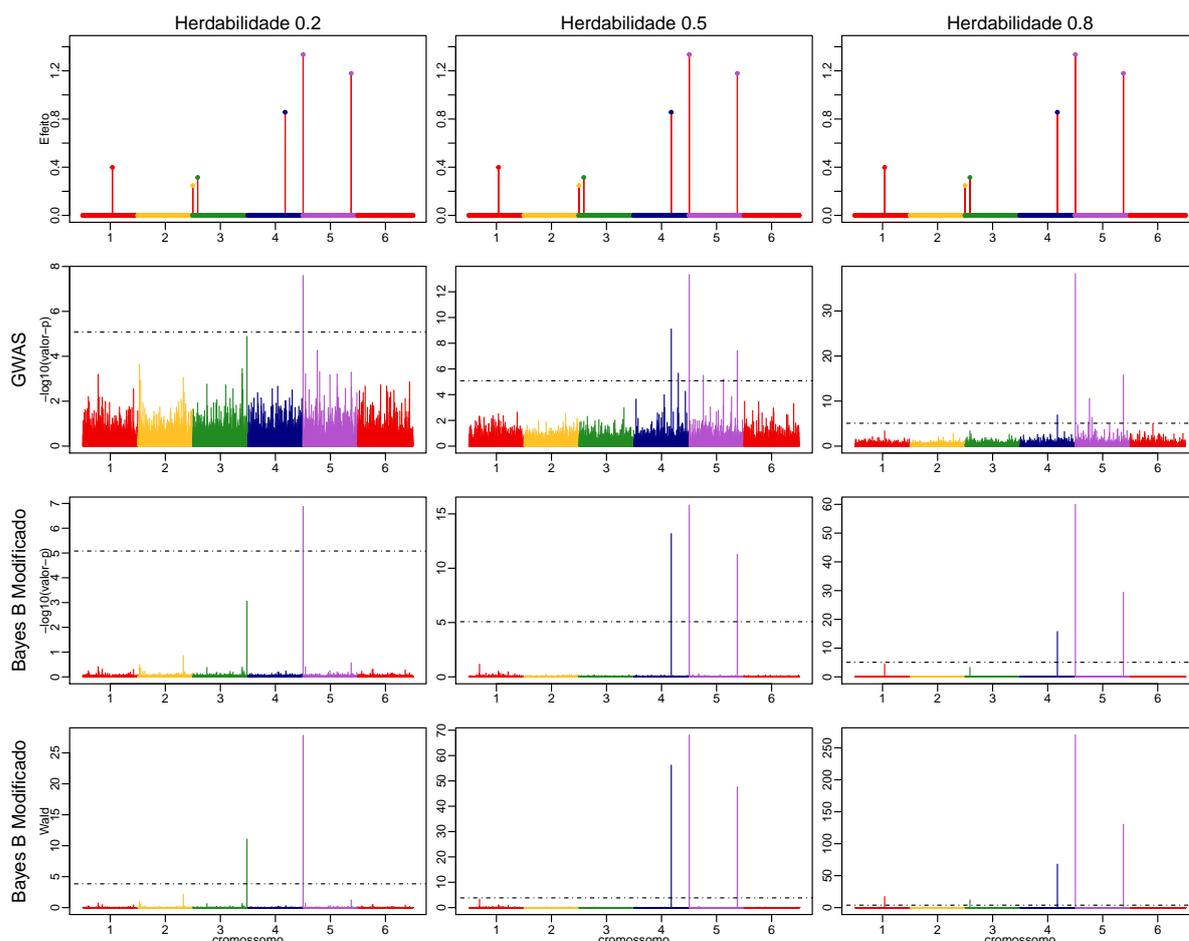
Na Figura 4.19, para o coeficiente de herdabilidade 0,8, pelo método GWAS, o painel usufrui de cinco associações significativas pelas marcas (5826; 3480; 5954; 1562 e 5033), em que a marca 5954 é uma associação falso positiva, localizada no cromossomo 6 e que percebe-se uma correlação com o SNP significativo de marca 5826. Para a mesma Figura, com o método Bayes A modificado, 10 SNPs foram associados pelo *valor - p* de Wald de forma significativa, dos quais 5 localizados no cromossomo 4, dentre eles, um SNP falso positivo (marca 3517). Isto confirma a interpretação da análise, no sentido de que existem genes nesta região com alta correlação, como também pode ser ao desequilíbrio de ligação LD. Alternativamente, como o cenário poligênico acusa mais de um gene significativo e, nesta região em especial, percebeu-se

uma concentração dos efeitos, um SNP diferente dos simulados rastreando a variante causal dos genes, indica a provável existência de mais SNPs significativos.

Alternativamente, como neste cenário poligênico, a concentração de efeitos significativos é maior, no caso especial da região localizada no cromossomo 4, um SNP diferente dos simulados conseguiu rastrear uma variante causal do gene, o que o torna um falso positivo, porém indica a provável existência de SNPs significativos ao redor, como mostra a Figura 4.19. Para o caso, o SNP cujo efeito significativo acusa maior proximidade seria a marca 3529.

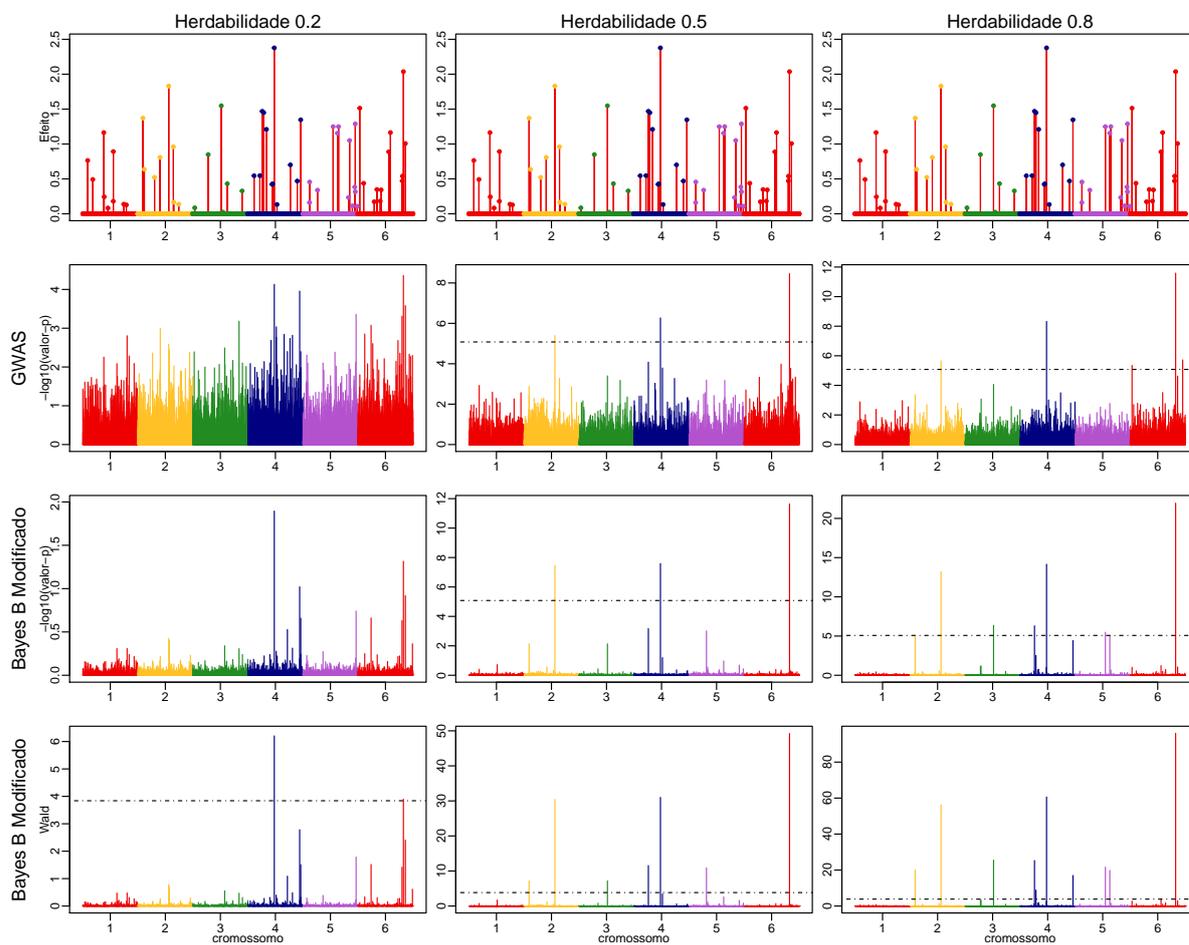
4.1.3.7 Análise Comparativa Bayes B modificado versus GWAS

Figura 4.20 – Cenário Oligogênico - Painéis demonstrando a análise dos efeitos dos marcadores, para cada uma das herdabilidades $h^2 \in \{0,2;0,5;0,8\}$, confrontando o método GWAS (Manhattan plot de $-\log_{10}(\text{valor} - p)$), com o método Bayes B modificado (Manhattan plot de $-\log_{10}(\text{valor} - p)$ e $\text{valor} - p$ de teste Wald), para associação de 6000 SNPs com total de 6 cromossomos



Fonte: Da autora (2019)

Figura 4.21 – Cenário Poligênico - Painéis demonstrando a análise dos efeitos dos marcadores, para cada uma das herdabilidades $h^2 \in \{0,2;0,5;0,8\}$, confrontando o método GWAS (Manhattan plot de $-\log_{10}(\text{valor} - p)$), com o método Bayes B modificado (Manhattan plot de $-\log_{10}(\text{valor} - p)$ e $\text{valor} - p$ de teste Wald), para associação de 6000 SNPs com total de 6 cromossomos



Fonte: Da autora (2019)

Tabela 4.9 – Detecção das marcas significativas por cromossomo. Resultados de associação genômica GWAS pacote rrBLUP e Método Bayes B modificado

| GWAS | | | | | cenário Poligênico | | | | | |
|---------------------|-------|------------|-------------------------|-------------------------|--------------------|-------|-------------|-------------------------|-------------------------|--|
| cenário Oligogênico | | | | | cenário Poligênico | | | | | |
| h^2 | Marca | Cr | $-\log_{10}(valor - p)$ | | h^2 | Marca | Cr | $-\log_{10}(valor - p)$ | | |
| 0,2 | 4004 | 5 | 7,597231* | | 0,2 | - | - | - | | |
| | 4004 | 5 | 13,339721* | | | 5826 | 6 | 8,459814* | | |
| | 3677 | 4 | 9,102989* | | | 0,5 | 3480 | 4 | 6,266500* | |
| | 4877 | 5 | 7,418931* | | | | 1562 | 2 | 5,390297* | |
| | 0,5 | 3805 | 4 | 5,668896** | | | 5826 | 6 | 11,58792* | |
| | | 4260 | 5 | 5,507352** | | | 3480 | 4 | 8,31891* | |
| 4625 | | 5 | 5,215285** | | 0,8 | 5954 | 6 | 5,72248** | | |
| 4004 | 5 | 38,278105* | | 1562 | | 2 | 5,66249* | | | |
| 4877 | 5 | 15,782135* | | 5033 | | 6 | 5,35015* | | | |
| 0,8 | 4260 | 5 | 10,589633** | | | | | | | |
| | 3677 | 4 | 6,920932* | | | | | | | |
| | 4303 | 5 | 6,386065** | | | | | | | |
| | 4237 | 5 | 5,372977** | | | | | | | |
| | | | | | | | | | | |
| Bayes B modificado | | | | | Cenário Poligênico | | | | | |
| cenário Oligogênico | | | | | Cenário Poligênico | | | | | |
| h^2 | Marca | Cr | Wald | $-\log_{10}(valor - p)$ | h^2 | Marca | Cr | Wald | $-\log_{10}(valor - p)$ | |
| 0,2 | 4004 | 5 | 27,8304760* | 6,8780345* | 0,2 | 3480 | 4 | 6,20564727* | 1,8950236 | |
| | 2983 | 3 | 11,06416008** | | | 5826 | 6 | 3,89694552* | 1,3153882 | |
| 0,5 | 4004 | 5 | 68,12398818* | 15,8138025* | | 5826 | 6 | 49,24104115* | 11,6451911* | |
| | 3677 | 4 | 56,20035255* | 13,1840917* | | 3480 | 4 | 31,01340838* | 7,5913327* | |
| | 4877 | 5 | 47,62465989* | 11,2872231* | | 1562 | 2 | 30,44027772* | 7,4630572* | |
| 0,8 | 4004 | 5 | 270,4021871* | 60,03274544* | 0,5 | 3261 | 4 | 11,54759547* | | |
| | 4877 | 5 | 130,3277279* | 29,45915560* | | 4317 | 5 | 10,88203386** | | |
| | 3677 | 4 | 67,90243253* | 15,76500438* | 2517 | 3 | 7,20009637* | | | |
| | 538 | 1 | 17,26160358* | | 1093 | 2 | 7,19851313* | | | |
| | 2088 | 3 | 12,17100162* | | 5826 | 6 | 95,9968557* | 21,9390527* | | |
| | | | | | 3480 | 4 | 60,6122248* | 14,1580086* | | |
| | | | | | 1562 | 2 | 56,2670242* | 13,1988183* | | |
| | | | | | 2517 | 3 | 25,5207743* | 6,3588778* | | |
| | | | | | 3261 | 4 | 25,2759594* | 6,3037632* | | |
| | | | | | 0,8 | 4549 | 5 | 21,6306717* | 5,4807972* | |
| | | | | 1093 | | 2 | 19,9307943* | 5,0953033* | | |
| | | | | 4629 | | 5 | 19,8147830* | | | |
| | | | | 3960 | | 4 | 16,9790909* | | | |
| | | | | 3288 | | 4 | 8,8439096* | | | |

* valor verdadeiro positivo; ** falso positivo; Cr=cromossomo

Fonte: Da autora (2019)

Para a análise dos métodos Bayes B modificado e GWAS, nos cenários oligogênico e poligênico, foram utilizados os mesmos coeficientes de herdabilidades 0,2, 0,5 e 0,8 para 6000 mil marcadores simulados. Na primeira coluna da Figura 4.20, em que se tratando da herdabilidade 0,2, apresentou-se o pico mais elevado, em ambos os gráficos relacionados a cada uma dos métodos. A marca 4004, com valor significativo para $-\log_{10}(valor - p) > 5,079$, pelos dois métodos e a marca 2983 como falso positivo, para o método Bayes B modificado. No cenário oligogênico, herdabilidade 0,5, verificaram-se três SNPs de $valor - p$ significativo e $\log_{10}(valor - p)$ de Wald para o método Bayes B modificado, contra seis efeitos significativos para GWAS, dos quais três marcas resultaram em falsos positivos (marcas 3805; 4260; 4625). Notavelmente, no mesmo painel de cenário oligogênico, para herdabilidade 0,8, foram detectados pela abordagem Bayes B modificado cinco $valores - p$ de Wald significativos em associação genômica (marcas 4004; 4877; 3677; 2088 e 538). Para a mesma análise, usando o

$-\log_{10}(\text{valor} - p) > 5,07$ de Wald, foram capturados somente três marcas (marca 3677, 4004 e 4877). Contrapondo com GWAS, o gráfico Manhattan apresenta um total de oito marcas significativas, observando-se cinco marcas falso positivas, localizadas nos cromossomos 5 e 6 (marcas 4260; 4303; 4237; 5412 e 4625).

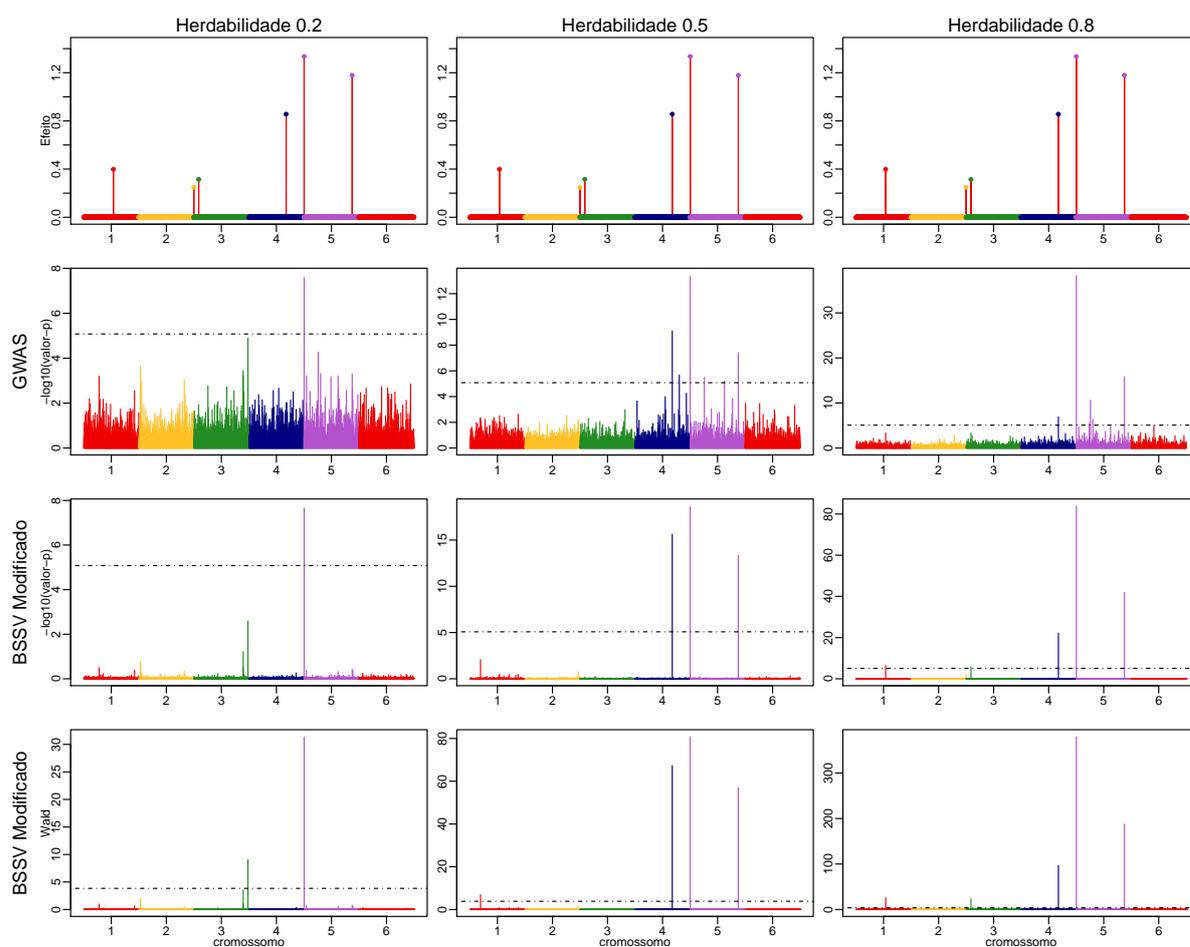
Para o cenário Poligênico, herdabilidade 0,2, método Bayes B modificado, sobressaíram duas marcas de $\text{valor} - p > 3,84$ de Wald significativo. O mapeamento associativo para GWAS não apresentou marcas com $-\log_{10}(\text{valor} - p)$ significativo (ver Tabela 4.9).

Os resultados gráficos, considerando a Figura 4.21, herdabilidade 0,5 para o método Bayes B modificado, capturaram sete marcas com $\text{valor} - p > 3,84$ de Wald significativo, sendo a marca 4317 falso positivo dentre as marcas 5826; 3480; 1562; 3261; 2517; 1093. Reduzindo este total para três marcas, quando calculadas pelo $-\log_{10}(\text{valor} - p)$ de Wald, também, foram constatadas as mesmas três marcas por GWAS (marcas 5826; 3480 e 1562).

Para a Figura 4.21, considerado a herdabilidade 0,8, o mapeamento de associação pode precisar os resultados da análise com uma quantidade de 10 SNPs de $\text{valores} - p$ de Wald significativos e, ainda não incluíram falso positivos, contra cinco SNPs por GWAS e ainda um SNP falso positivo (marca 5954).

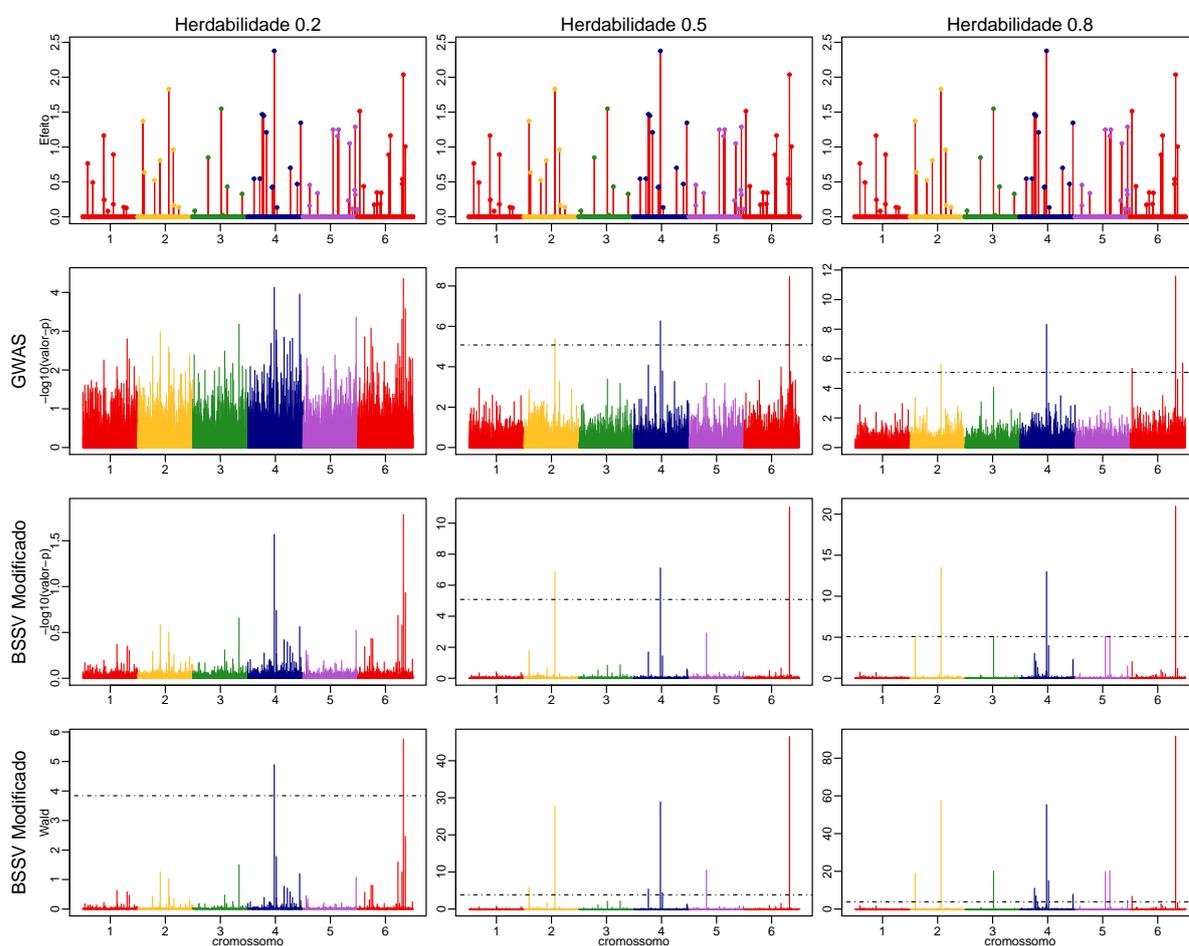
4.1.3.8 Análise Comparativa BSSV modificado versus GWAS

Figura 4.22 – Cenário Oligogênico - Painéis demonstrando a análise dos efeitos dos marcadores, para cada uma das herdabilidades $h^2 \in \{0,2;0,5;0,8\}$, confrontando o método GWAS (Manhattan plot de $-\log_{10}(\text{valor} - p)$), com o método BSSV modificado (Manhattan plot de $-\log_{10}(\text{valor} - p)$ e $\text{valor} - p$ de teste Wald), para associação de 6000 SNPs com total de 6 cromossomos



Fonte: Da autora (2019)

Figura 4.23 – Cenário Poligênico - Painéis demonstrando a análise dos efeitos dos marcadores, para cada uma das herdabilidades $h^2 \in \{0,2;0,5;0,8\}$, confrontando o método GWAS (Manhattan plot de $-\log_{10}(\text{valor} - p)$), com o método BSSV modificado (Manhattan plot de $-\log_{10}(\text{valor} - p)$ e $\text{valor} - p$ de teste Wald), para associação de 6000 SNPs com total de 6 cromossomos



Fonte: Da autora (2019)

Tabela 4.10 – Detecção das marcas significativas por cromossomo. Resultados de associação genômica GWAS pacote rrBLUP e Método BSSV modificado

| GWAS | | | | | BSSV modificado | | | | | | |
|---------------------|-------|------|--------------------------------|--------------------|-----------------|------|--------------------------------|-----------|--------------------------------|--|--|
| cenário Oligogênico | | | | cenário Poligênico | | | | | | | |
| h^2 | Marca | Cr | $-\log_{10}(\text{valor} - p)$ | h^2 | Marca | Cr | $-\log_{10}(\text{valor} - p)$ | | $-\log_{10}(\text{valor} - p)$ | | |
| 0,2 | 4004 | 5 | 7,597231* | 0,2 | - | - | - | | | | |
| | 4004 | 5 | 13,339721* | | 0,5 | 5826 | 6 | 8,459814* | | | |
| | 3677 | 4 | 9,102989* | | | 3480 | 4 | 6,266500* | | | |
| | 4877 | 5 | 7,418931* | | | 1562 | 2 | 5,390297* | | | |
| | 0,5 | 3805 | 4 | | 5,668896** | 0,8 | 5826 | 6 | 11,58792* | | |
| | | 4260 | 5 | | 5,507352** | | 3480 | 4 | 8,31891* | | |
| 4625 | | 5 | 5,215285** | 5954 | 6 | | 5,72248** | | | | |
| 4004 | | 5 | 38,278105* | 1562 | 2 | | 5,66249* | | | | |
| 0,8 | 4877 | 5 | 15,782135* | 0,8 | 5033 | 6 | 5,35015* | | | | |
| | 4260 | 5 | 10,589633** | | | | | | | | |
| | 3677 | 4 | 6,920932* | | | | | | | | |
| | 4303 | 5 | 6,386065** | | | | | | | | |
| | 4237 | 5 | 5,372977** | | | | | | | | |
| | | | | | | | | | | | |

* valor verdadeiro positivo; ** falso positivo; Cr=cromossomo

Fonte: Da autora (2019)

Para o método BSSV modificado, no cenário oligogênico, com coeficiente de herdabilidade 0,2, apresentado pela Figura 4.22, dois SNPs de marca 4004 e 2983 foram associados de forma a observar o $\text{valor} - p$ de Wald maior que 3,84. Destes dois SNPs, o de marca 2983 é considerado um falso positivo. O modelo em que se baseou o coeficiente de herdabilidade 0,5, pelo método BSSV, capturou quatro SNPs associados ao gene simulado, com $\text{valor} - p$ de Wald maior que 3,84, localizados no cromossomos 1, 4 e 5, cujas marcas são 193, 3677, 4004 e 4877. A marca 193 caracteriza-se como um falso positivo. Ao confrontar com GWAS, que apresenta seis marcas significativas pelo $-\log_{10}(\text{valor} - p)$ ser igual, percebem-se três SNPs falsos positivos marcas 3805, 4260, 4625 (ver Tabela 4.10). Neste caso, ainda assim, é possível confiar, de maneira geral, por meio da análise gráfica, no método BSSV modificado. Pelos resultados

de comparação entre as significâncias dos marcadores SNPs, BSSV modificado apresenta nível de confiança, provavelmente maior do que o nível de confiança exibido por GWAS.

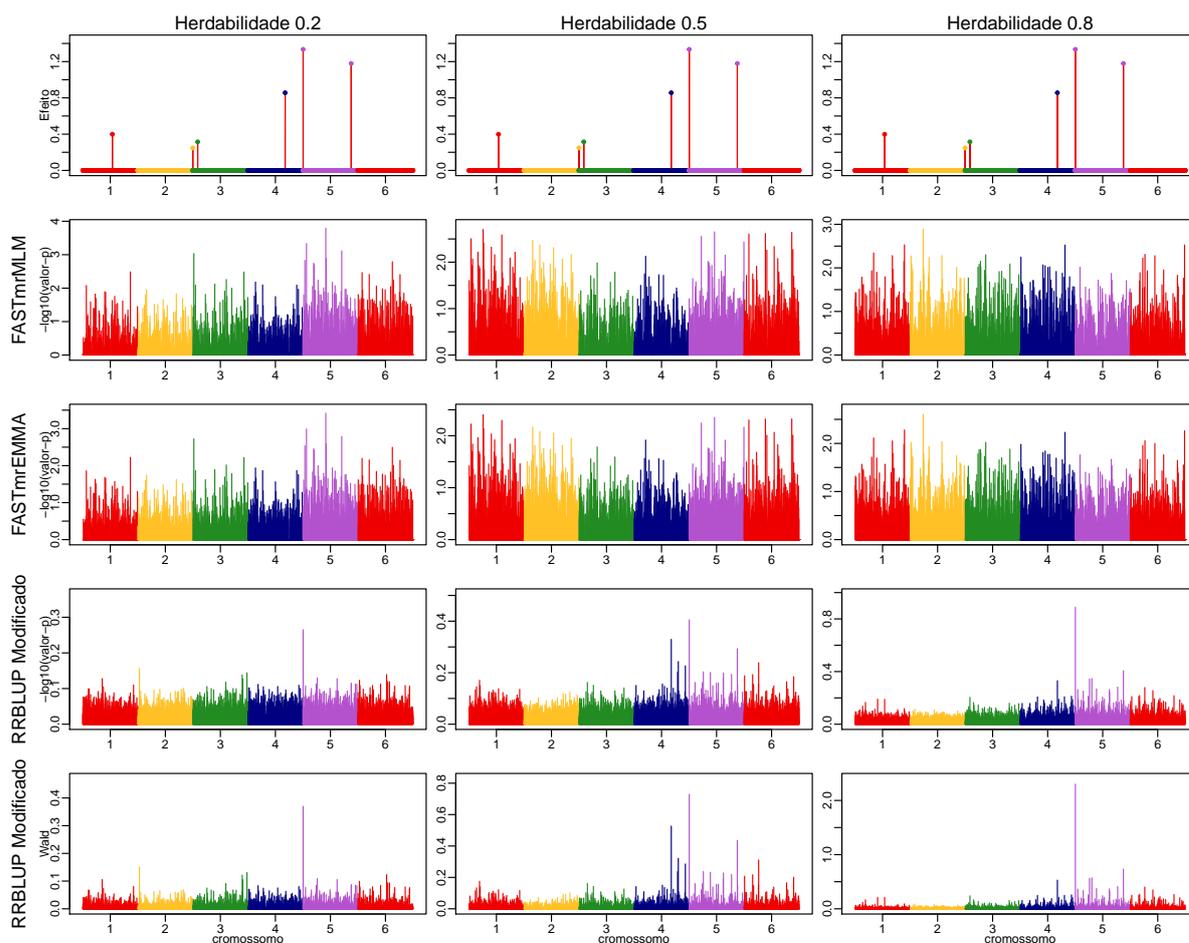
Neste mesmo painel da Figura 4.22, quando se trata da herdabilidade 0,8, pelo método BSSV modificado, tem-se uma detecção completa dos SNPs significativos pelo *valor - p* e $-\log_{10}(\text{valor} - p)$ de Wald para os cinco marcadores: 4004, 4877, 3677, 2088 e 538, contra três significativos e cinco falso positivos por GWAS. (Tabela 4.10).

A análise observada na Figura 4.23 demonstrou um aumento do número de elementos significativos, com o aumento da herdabilidade. Destaca-se o método BSSV modificado, que associou duas marcas (5826 e 3480) pelo *valor - p* de Wald e nenhuma pelo $-\log_{10}(\text{valor} - p)$, quando coeficiente de herdabilidade assume valor 0,2. Seria conveniente, neste sentido de escolha do teste, optar pelo teste com valor de Wald, uma vez que, realizando a transformação logarítmica na base 10, o número de falso negativo aumenta. Por GWAS, nenhuma associação foi significativa.

Para os painéis de coeficiente de herdabilidade 0,5, verificaram-se cinco associações SNPs significativos pelo valor-p de Wald e dois falsos positivos (marcas 4317 e 3517) pelo método BSSV modificado, contrapondo com 3 SNPs de associação genômica significativa com $-\log_{10}(\text{valor} - p) > 5,07$. Já no cenário, em que a herdabilidade assume valor 0,8, quatorze foram as associações significativas pelo método BSSV modificado, das quais dois efeitos são falso positivos como visto pela Tabela 4.10. Ao aplicar a transformação logarítmica na base 10, reduziu para um total de seis associações significativas (marcas 5826; 1562; 3480; 4629; 2517 e 4549). Ainda assim, contrapondo à GWAS, que fez associação genômica de cinco marcas, sendo uma falso positivo (marca 5954).

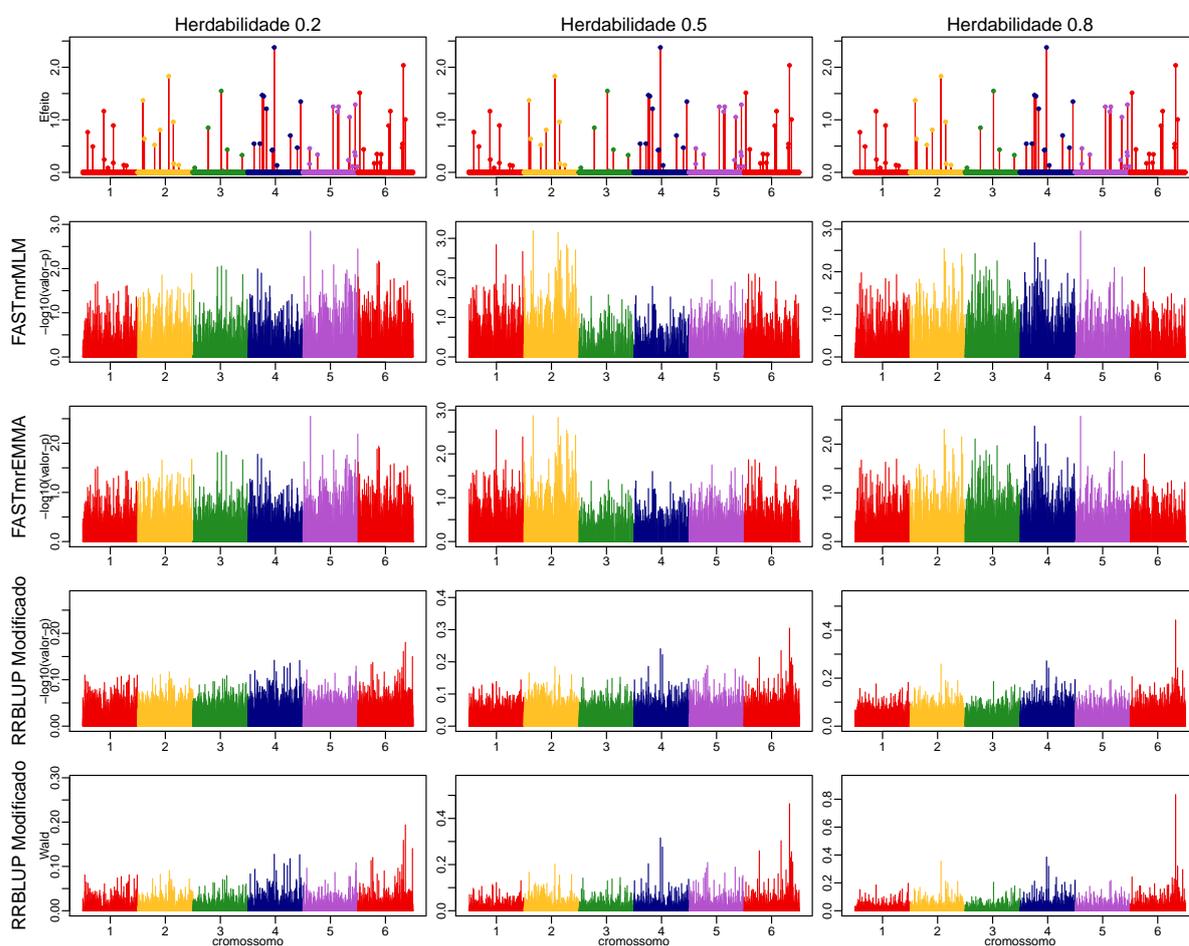
4.1.3.9 Análise Comparativa RR-BLUP modificado versus FASTmrEMMA FASTmrMLM

Figura 4.24 – Cenário Oligogênico - Painéis demonstrando a análise dos efeitos dos marcadores, para cada uma das herdabilidades $h^2 \in \{0, 2; 0, 5; 0, 8\}$, confrontando o método FASTmrEMMA e FASTmrMLM (Manhattan plot de $-\log_{10}(\text{valor} - p)$), com o método RR-BLUP modificado (Manhattan plot de $-\log_{10}(\text{valor} - p)$ e $\text{valor} - p$ do teste Wald), para associação de 6000 SNPs com total de 6 cromossomos



Fonte: Da autora (2019)

Figura 4.25 – Cenário Poligênico - Painéis demonstrando a análise dos efeitos dos marcadores, para cada uma das herdabilidades $h^2 \in \{0,2;0,5;0,8\}$, confrontando o método FASTmrEMMA e FASTmrMLM (Manhattan plot de $-\log_{10}(\text{valor} - p)$), com o método RR-BLUP modificado (Manhattan plot de $-\log_{10}(\text{valor} - p)$ e $\text{valor} - p$ de teste Wald), para associação de 6000 SNPs com total de 6 cromossomos

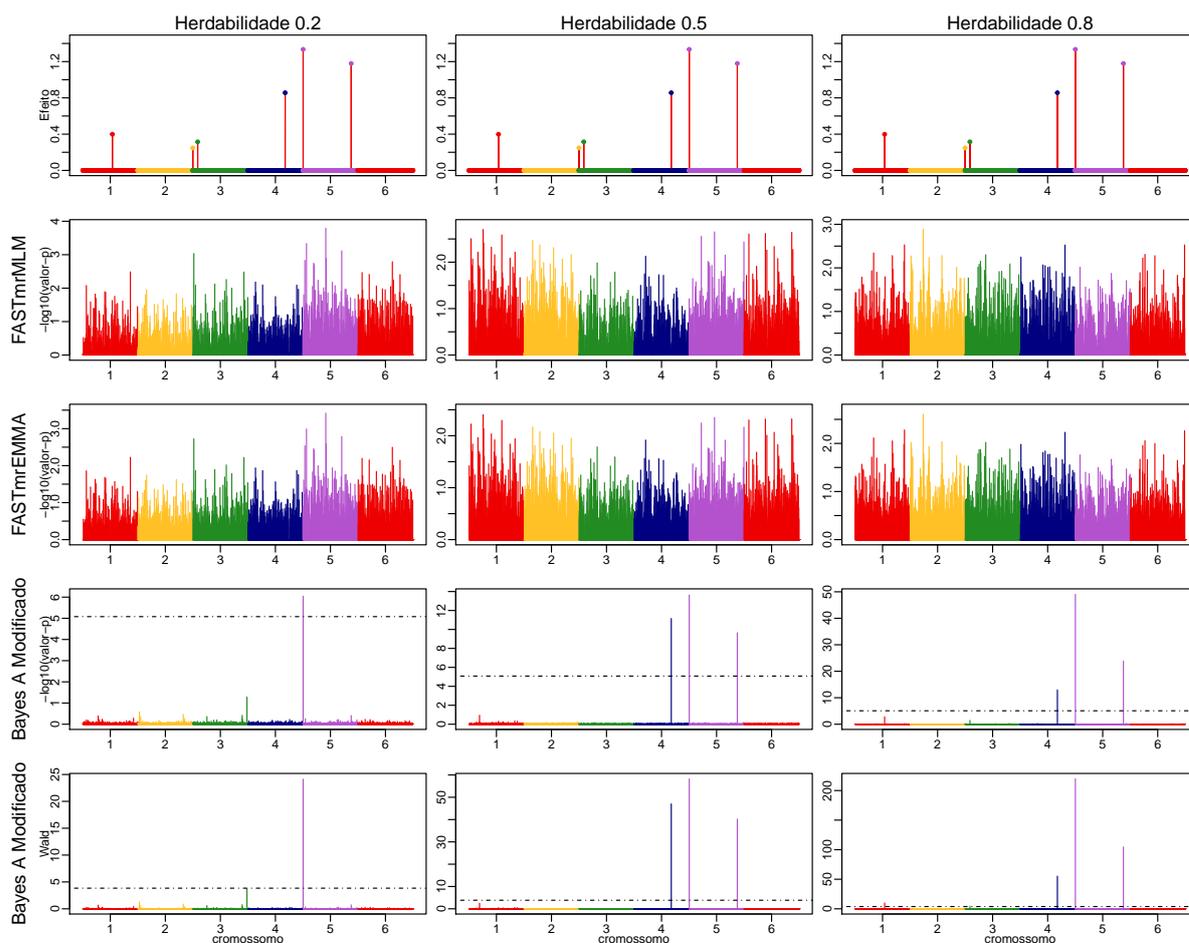


Fonte: Da autora (2019)

Utilizando dois métodos GWAS multiloci denominados FASTmrMLM e FASTmrEMMA, pelo pacote mrMLM, não se identificou nenhum gene significativo para a característica simulada, baseando-se nos gráficos das Figuras 4.24 4.25, em que a limiar de logaritmo de probabilidades $-\log_{10}(\text{valor} - p) > 5,07$. Ainda apresentaram comportamento bem distintos quando comparados ao método RR-BLUP modificado. Uma comparação entre o método FASTmrMLM e FASTmrEMMA é que, embora não tenham estimativas de efeito significativo, aparentam a mesma disposição gráfica, a menos de uma magnitude (ver Tabelas 50 a 53, em anexo).

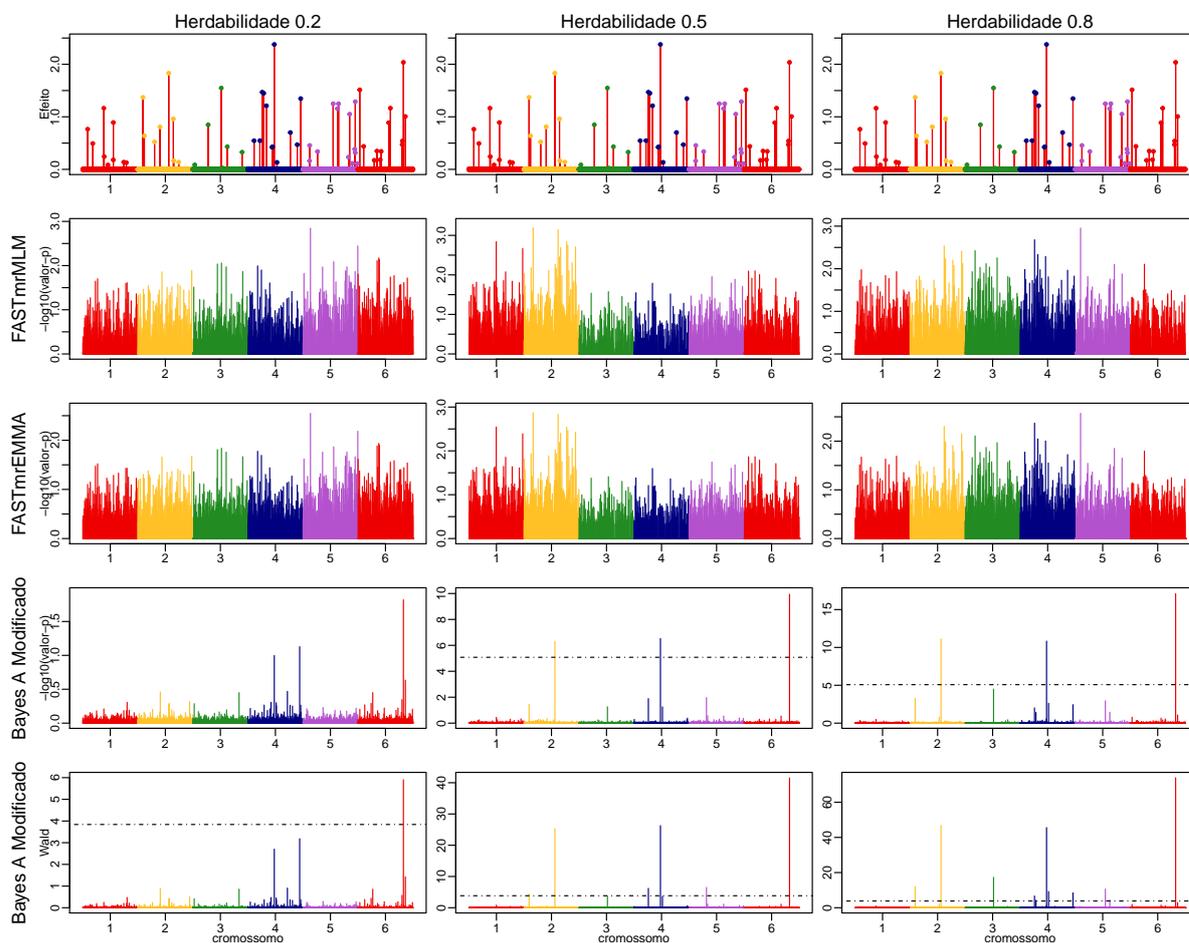
4.1.3.10 Análise Comparativa Bayes A modificado versus FASTmrEMMA e FASTmrMLM

Figura 4.26 – Cenário Oligogênico - Painéis demonstrando a análise dos efeitos dos marcadores, para cada uma das herdabilidades $h^2 \in \{0, 2; 0, 5; 0, 8\}$, confrontando o método FASTmrEMMA e FASTmrMLM (Manhattan plot de $-\log_{10}(\text{valor} - p)$), com o método Bayes A modificado (Manhattan plot de $-\log_{10}(\text{valor} - p)$ e $\text{valor} - p$ do teste Wald), para associação de 6000 SNPs com total de 6 cromossomos



Fonte: Da autora (2019)

Figura 4.27 – Cenário Poligênico - Painéis demonstrando a análise dos efeitos dos marcadores, para cada uma das herdabilidades $h^2 \in \{0,2; 0,5; 0,8\}$, confrontando o método FASTmrEMMA e FASTmrMLM (Manhattan plot de $-\log_{10}(\text{valor} - p)$), com o método Bayes A modificado (Manhattan plot de $-\log_{10}(\text{valor} - p)$ e $\text{valor} - p$ de teste Wald), para associação de 6000 SNPs com total de 6 cromossomos

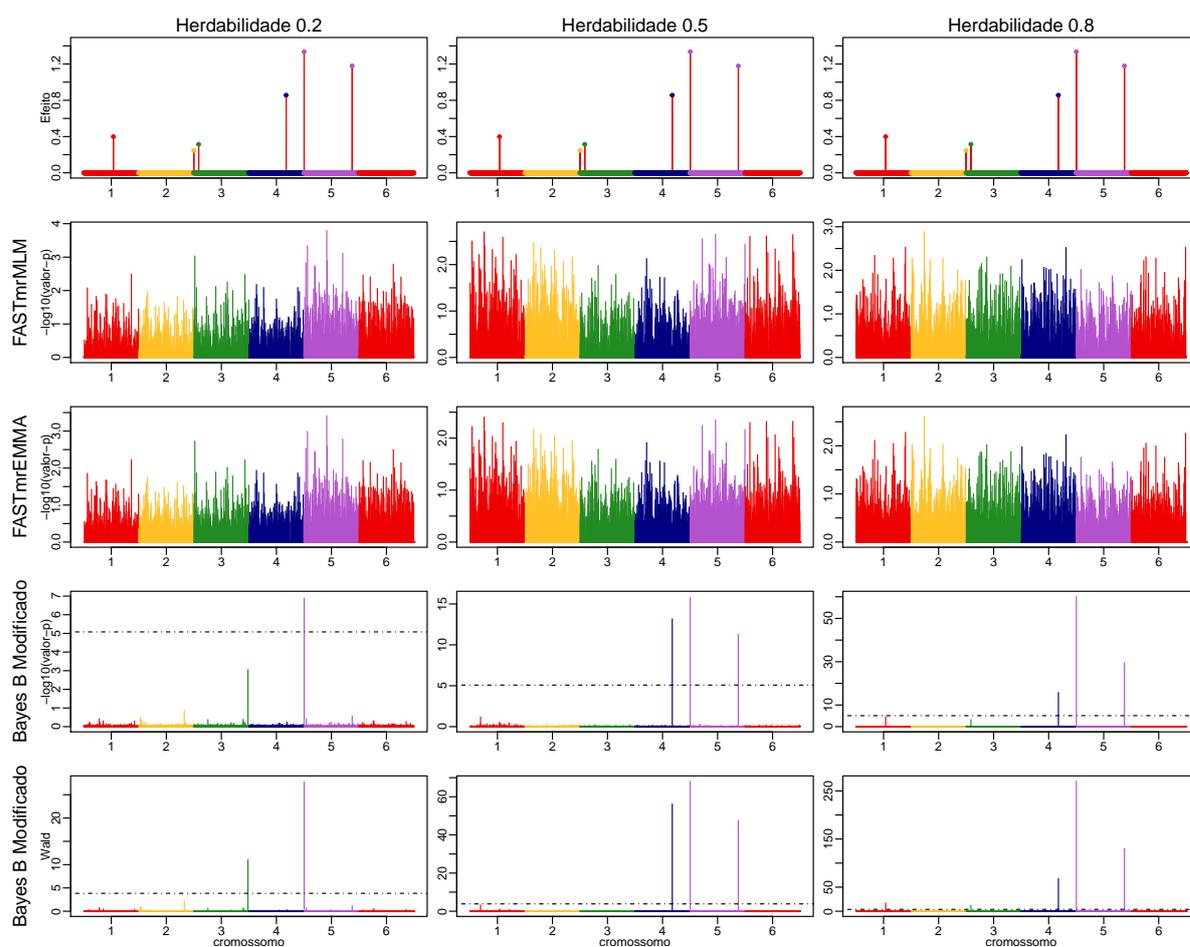


Fonte: Da autora (2019)

Comparando os resultados das análises bayesianas pelo método Bayes A modificado e os métodos FASTmrEMMA e FASTmrMLM, para os cenários oligogênico e poligênico e, considerando os três coeficientes de herdabilidade, nenhum efeito significativo, para os métodos FASTmrEMMA e FASTmrMLM, pode-se afirmar que foram melhores resultados em associação obtidos pelo método Bayes A modificado (Figuras 4.26 e 4.27).

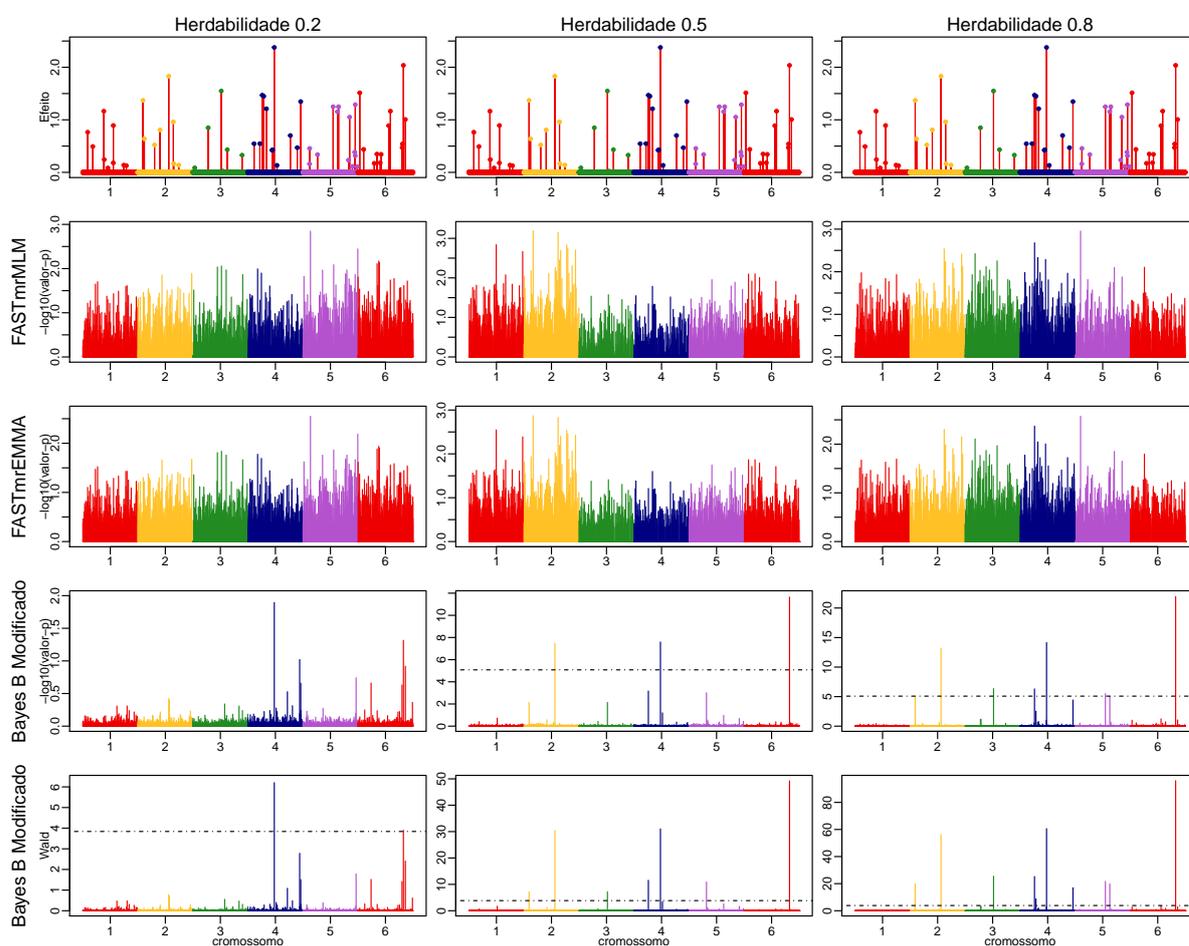
4.1.3.11 Análise Comparativa Bayes B modificado versus FASTmrEMMA FASTmrMLM

Figura 4.28 – Cenário Oligogênico - Painéis demonstrando a análise dos efeitos dos marcadores, para cada uma das herdabilidades $h^2 \in \{0,2; 0,5; 0,8\}$, confrontando o método FASTmrEMMA e FASTmrMLM (Manhattan plot de $-\log_{10}(\text{valor} - p)$), com o método Bayes B modificado (Manhattan plot de $-\log_{10}(\text{valor} - p)$ e $\text{valor} - p$ do teste Wald), para associação de 6000 SNPs com total de 6 cromossomos



Fonte: Da autora (2019)

Figura 4.29 – Cenário Poligênico - Painéis demonstrando a análise dos efeitos dos marcadores, para cada uma das herdabilidades $h^2 \in \{0,2;0,5;0,8\}$, confrontando o método FASTmrEMMA e FASTmrMLM (Manhattan plot de $-\log_{10}(\text{valor} - p)$), com o método Bayes B modificado (Manhattan plot de $-\log_{10}(\text{valor} - p)$ e $\text{valor} - p$ de teste Wald), para associação de 6000 SNPs com total de 6 cromossomos

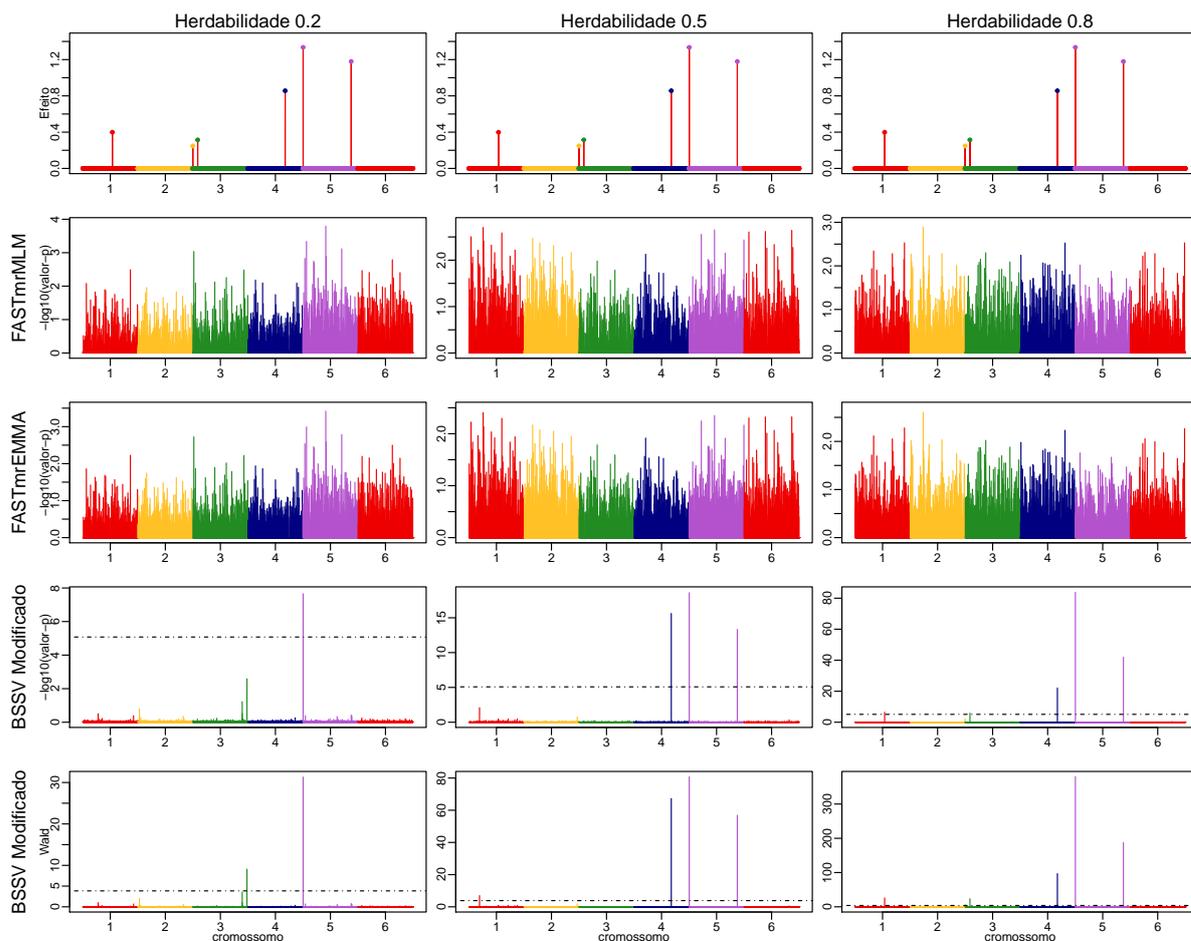


Fonte: Da autora (2019)

Os resultados gráficos, considerando as Figuras 4.8 e 4.9, para os métodos Bayes B modificado em comparação com os métodos FASTmrMLM e FASTmrEMMA, o método Bayes B modificado apresentou um resultado melhor, na detecção dos efeitos significativos, em comparação com os métodos FASTmrMLM e FASTmrEMMA.

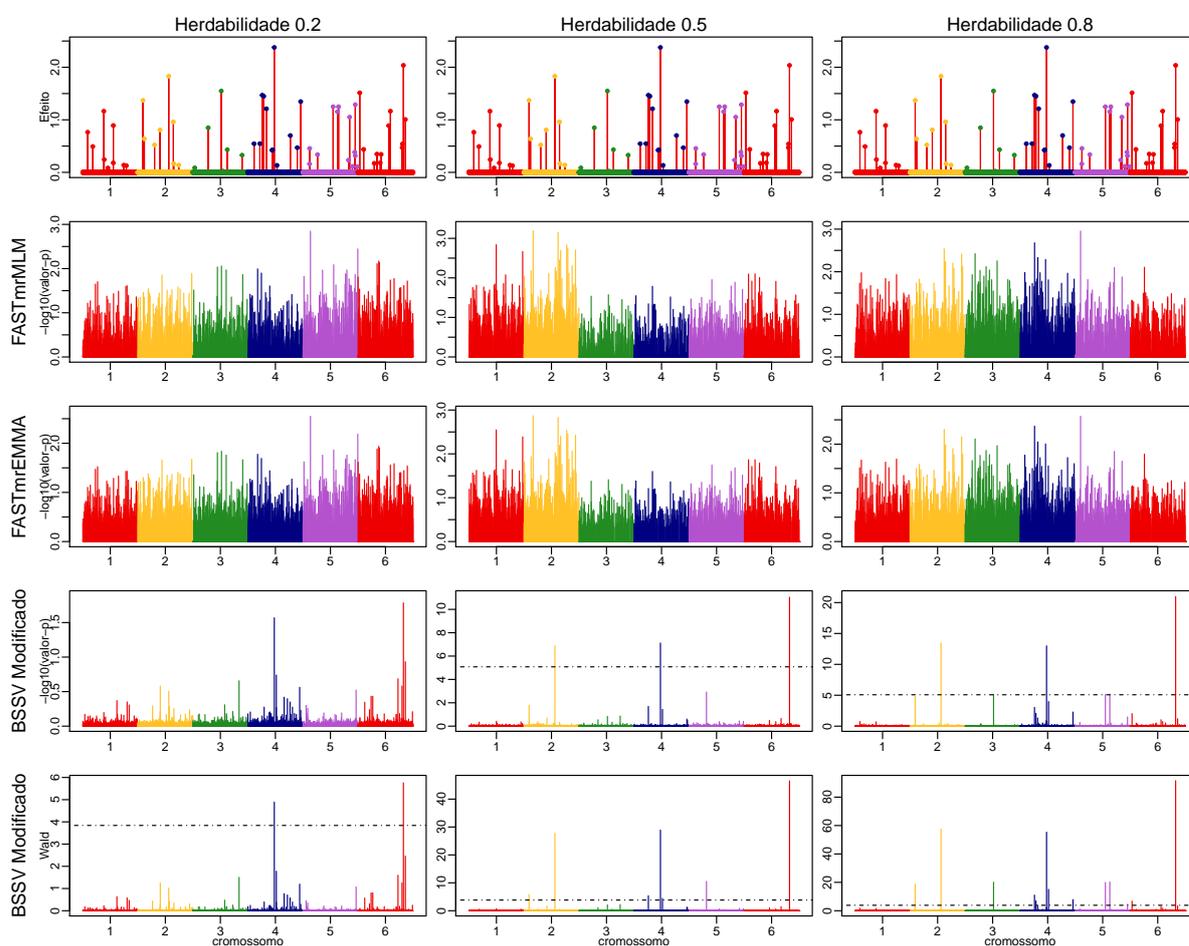
4.1.3.12 Análise Comparativa BSSV modificado versus FASTmrMLM e FASTmrEMMA

Figura 4.30 – Cenário Oligogênico - Painéis demonstrando a análise dos efeitos dos marcadores, para cada uma das herdabilidades $h^2 \in \{0, 2; 0, 5; 0, 8\}$, confrontando o método FASTmrEMMA e FASTmrMLM (Manhattan plot de $-\log_{10}(\text{valor} - p)$), com o método BSSV modificado (Manhattan plot de $-\log_{10}(\text{valor} - p)$ e $\text{valor} - p$ de teste Wald), para associação de 6000 SNPs com total de 6 cromossomos



Fonte: Da autora (2019)

Figura 4.31 – Cenário Poligênico - Painéis demonstrando a análise dos efeitos dos marcadores, para cada uma das herdabilidades $h^2 \in \{0,2;0,5;0,8\}$, confrontando o método FASTmrEMMA e FASTmrMLM (Manhattan plot de $-\log_{10}(\text{valor} - p)$), com o método BSSV modificado (Manhattan plot de $-\log_{10}(\text{valor} - p)$ e $\text{valor} - p$ de teste Wald), para associação de 6000 SNPs com total de 6 cromossomos



Fonte: Da autora (2019)

Para o método BSSV modificado, tanto no cenário oligogênico quanto no cenário poligênico (Figuras 4.31 e 4.31), comparados com os métodos FASTmrMLM e FASTmrEMMA, percebe-se que não ocorreu associação de SNPs significativos pelo $-\log_{10}(\text{valor} - p)$, para os métodos FASTmrMLM e FASTmrMLM, contrapondo com BSSV modificado, que capturou SNPs significativos pelo $\text{valor} - p$ de Wald em todos os painéis 4.10.

4.2 Comparação dos métodos propostos em aplicação a dados de resistência à podridão da espiga causada por *Fusarium verticillioides*

Pretendeu-se comparar os métodos propostos: RR-BLUP modificado, Bayes A modificado, Bayes B modificado e BSSV modificado, com os métodos RR-BLUP, Bayes A, Bayes B, BSSV, GWAS e FASTmrMLM, utilizando os dados do experimento de milho, avaliadas 230 linhagens em análise conjunta (locais Lavras e Uberlândia) para a proporção de espigas com sintomas de podridão. O material genético utilizado consistiu de 23153 marcadores do tipo SNPs, derivados da metodologia DART-seq (*Diversity Arrays Technology*), num total de 10 cromossomos identificados e, a partir da posição 11, remete como posição de cromossomos não identificados.

Os experimentos foram conduzidos, utilizando o delineamento experimental de blocos incompletos com três repetições por local e compostas por uma linha de 3 metros com espaçamento de 0,7 metros. Dados de podridão foram submetidos à análise, utilizando modelos mistos e aplicação aos métodos RR-BLUP, RR-BLUP modificado, Bayes A, Bayes A modificado, Bayes B, Bayes B modificado, BSSV, BSSV modificado, GWAS e FASTmrMLM.

Tabela 4.11 – Estimativas da variância genética aditiva (σ_a^2) e residual (σ_e^2) e estimativa da herdabilidade (h^2) para análise conjunta de Lavras e Uberlândia

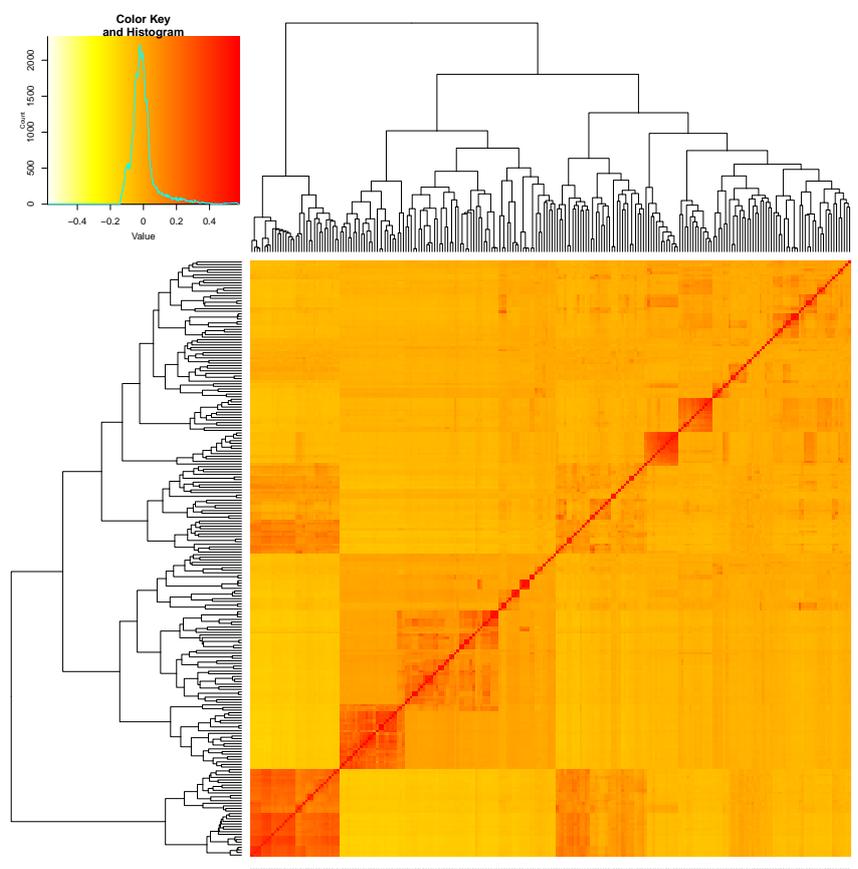
| | h^2 | σ_a^2 | σ_e^2 | $\sigma_a^2 + \sigma_e^2$ |
|------------------|-------|--------------|--------------|---------------------------|
| Análise conjunta | 0,056 | 0,0047 | 0,08 | 0,0847 |

Fonte: Da autora (2019)

Para as análises fenotípicas, para resistência à podridão de espiga causada por *Fusarium verticillioides*, foram obtidas as estimativas da variância genética aditiva, variância residual e a herdabilidade. A proporção da variância total explicada pela variância genética para a análise conjunta foi de 5,6% (Tabela 4.11).

É possível observar um agrupamento com definição mais nítida formada, na base inferior esquerda da Figura 4.32, sugerindo que, dentre todas as linhagens esse grupo apresentou maior parentesco genômico entre as linhagens a ele pertencente. Na parte mediana da Figura, observa-se uma sobreposição de grupos em tom alaranjado, em que se localizam dois subgrupos bem distintos em tons vermelhos, nos quais é possível observar que, em pequenos agrupamentos, ocorre um nível intermediário de parentesco. Na parte diagonal superior direito da Figura 4.32, também, verificam-se pequenos subgrupos em tons vermelhos, com parentesco alto.

Figura 4.32 – *Heatmap* e dendrograma das 230 linhagens da matriz de parentesco mostrando diferença na expressão gênica. Vermelho indica alta correlação entre os pares de marcadores e o amarelo, baixa correlação



Fonte: Da autora (2019)

Tabela 4.12 – Resumo dos resultados de seleção de modelo pelo critério de informação bayesiano (BIC)

| Modelo | Número de PCs/ Covariáveis | BIC | log da verossimilhança |
|----------|----------------------------|----------|------------------------|
| Conjunta | 0 | -51,3992 | -43,2421 |
| | 1 | -54,1528 | -43,2767 |
| | 2 | -56,0837 | -42,4885 |
| | 3 | -56,5804 | -40,2662 |

Fonte: Da autora (2019)

Pelo Critério de Informação Bayesiana (BIC), apresentado na Tabela 4.12, determinou-se o número de componentes principais necessários para explicar adequadamente a estrutura populacional. Para a análise conjunta, o número de componentes principais necessários foi zero. Todavia não há garantia de que o número de componentes para explicar a estrutura da população, seja realmente o melhor, uma vez que o grau de correlação com a estrutura populacional varia de característica para característica (LIPKA et al., 2012).

Tabela 4.13 – Estimativas do coeficiente de herdabilidade \hat{h}^2 , R^2 e PRESS para cada um dos métodos

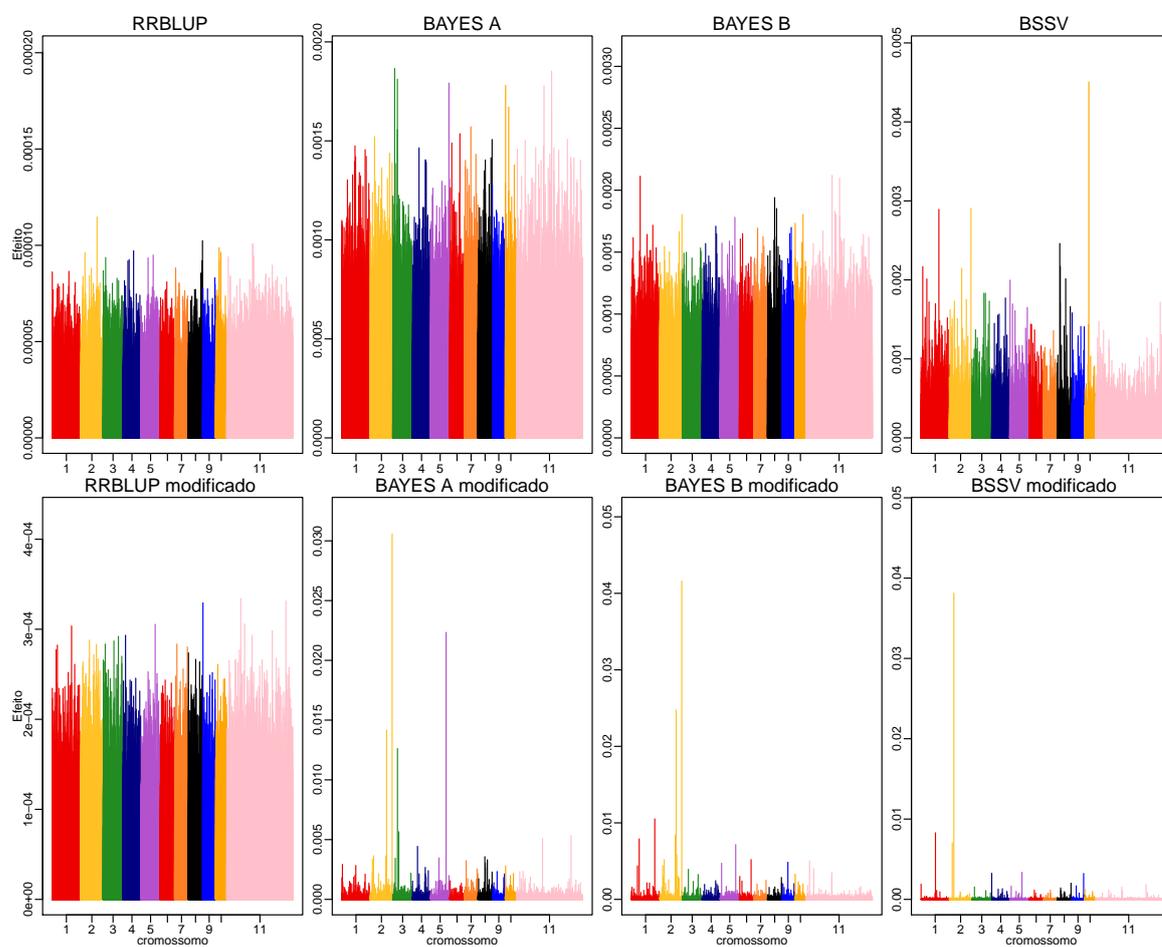
| | \hat{h}^2 | R^2 | PRESS | |
|---------|--------------------|-------|--------|--------|
| métodos | RR-BLUP | 0,06 | 0,3144 | 0,6818 |
| | RR-BLUP modificado | 0,13 | 0,4391 | 0,6730 |
| | Bayes A | 0,45 | 0,8735 | 0,6271 |
| | Bayes A modificado | 0,28 | 0,6778 | 0,6533 |
| | Bayes B | 0,25 | 0,7459 | 0,6440 |
| | Bayes B modificado | 0,25 | 0,6278 | 0,6575 |
| | BSSV | 0,14 | 0,6173 | 0,6701 |
| | BSSV modificado | 0,06 | 0,3598 | 0,6820 |

Fonte: Da autora (2019)

As estimativas dos efeitos de marcadores variaram, de acordo com a pressuposição do modelo genético, associada a cada metodologia de análise. Apresentaram comportamento distintos nas comparações entre o método e o respectivo método modificado e, portanto, para comparar as estatística PRESS e R^2 - calculadas pela observações obtidas, em cada modelo e valores estimados - é necessário examinar detalhadamente cada um dos métodos e sua relação de maior ou menor encolhimento do efeito estimado.

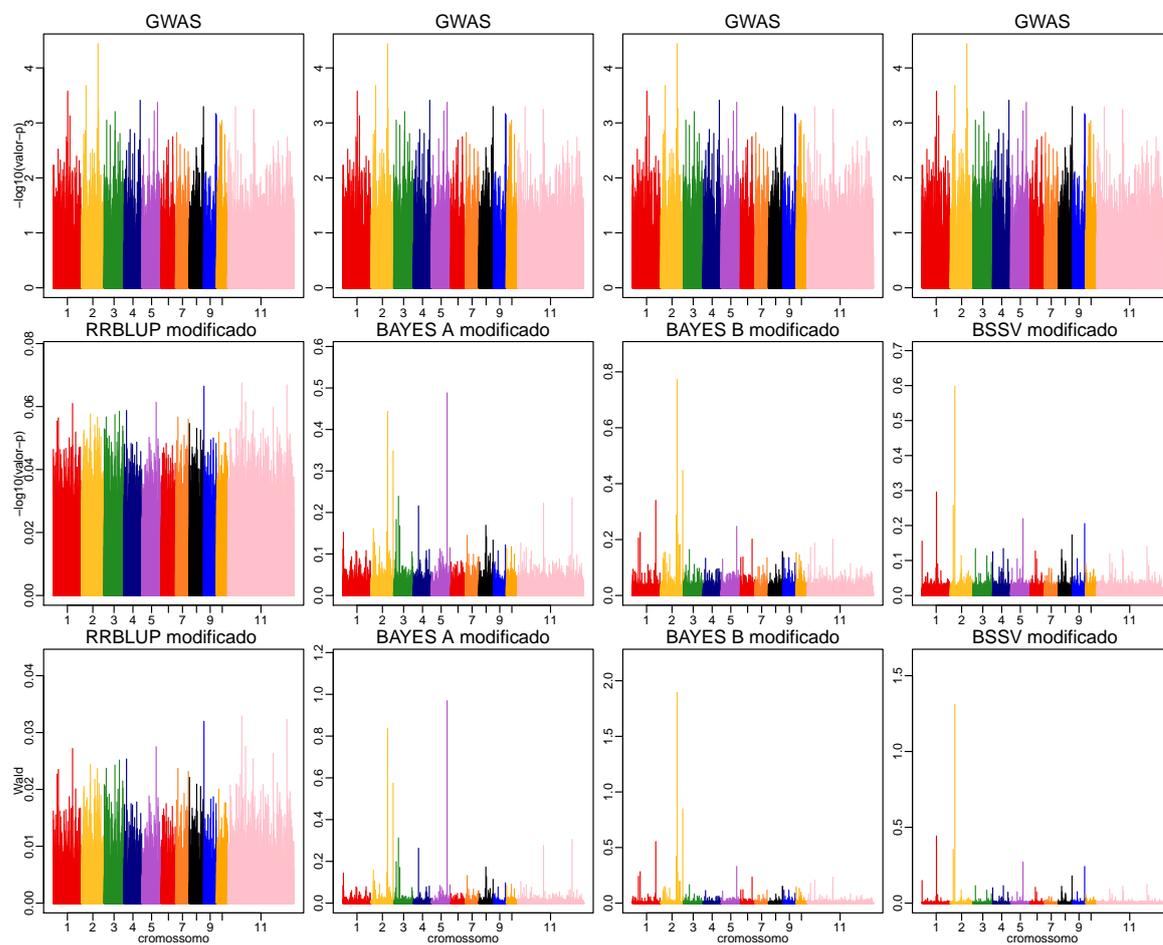
No caso dos métodos propostos, o que ocorre em comparação aos efeitos estimados, em cada método e seu respectivo método modificado, é um maior encolhimento absoluto para os métodos RR-BLUP, Bayes A, Bayes B e BSSV (Figura 4.33), o que gerou um aumento no valor da PRESS, mas que não determina uma confiabilidade, ao afirmar sobre a melhora do método em comparação ao modificado. Esse aumento do valor da PRESS, para os métodos clássicos, refere-se exatamente ao encolhimento, já que maior encolhimento gera um maior valor de PRESS. Neste caso, a análise pode ser confiada pelo apoio gráfico da Figura 4.35, principalmente, em comparação com o método FASTmrMLM, que identificou marcadores associados à característica de resistência à podridão da espiga e permitiu identificar soluções muito boas pelos métodos modificados, reafirmando a eficiência nos métodos modificados (RR-BLUP modificado, Bayes A modificado, Bayes B modificado e BSSV modificado). A análise realizada com o RR-BLUP e RR-BLUP modificado foi evidente em retratar o comportamento da característica de encolhimento e, ainda observa-se menor PRESS para o método RR-BLUP modificado (Tabela 4.13). Por outro lado, os métodos Bayes A e Bayes A modificado apresentaram menor encolhimento em comparação com Bayes A modificado e Bayes B modificado e foram com menores valores de PRESS (0,6271 e 0,6440).

Figura 4.33 – Painéis demonstrando a análise dos efeitos dos marcadores, confrontando os métodos RR-BLUP, Bayes A, Bayes B e BSSV, com os métodos RR-BLUP modificado, Bayes A modificado, Bayes B modificado e BSSV modificado, para associação de 23153 SNPs



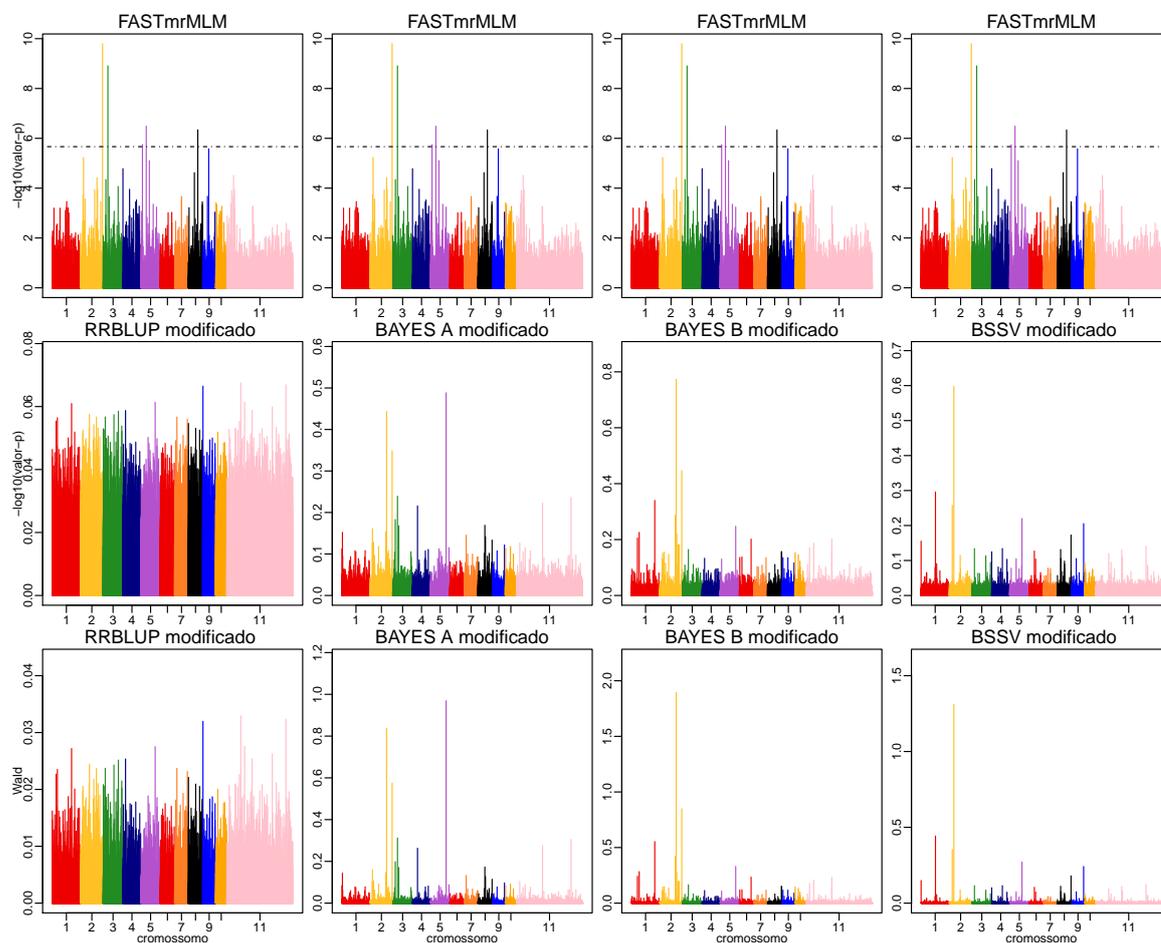
Fonte: Da autora (2019)

Figura 4.34 – Painéis demonstrando a análise dos efeitos dos marcadores, confrontando o método GWAS (Manhattan plot de $-\log_{10}(\text{valor} - p)$), com os métodos RR-BLUP modificado, Bayes A modificado, Bayes B modificado e BSSV modificado (Manhattan plot de $-\log_{10}(\text{valor} - p)$ e $\text{valor} - p$ de teste Wald), para associação de 23153 SNPs



Fonte: Da autora (2019)

Figura 4.35 – Painéis demonstrando a análise dos efeitos dos marcadores, confrontando o método FASTmrMLM (Manhattan plot de $-\log_{10}(\text{valor} - p)$), com os métodos RR-BLUP modificado, Bayes A modificado, Bayes B modificado e BSSV modificado (Manhattan plot de $-\log_{10}(\text{valor} - p)$ e $\text{valor} - p$ de teste Wald), para associação de 23153 SNPs



Fonte: Da autora (2019)

Tabela 4.14 – Detecção das marcas significativas por cromossomo. Resultados de associação genômica FASTmrMLM pacote mrMLM, para associação de 23153 SNPs

| FASTmrMLM | | | |
|-----------|------------|-----------|--------------------------------|
| nome | cromossomo | Posição | $-\log_{10}(\text{valor} - p)$ |
| 4588524 | 2 | 234586529 | 9,797536 |
| 2455066 | 3 | 43228304 | 8,906851 |
| 4766933 | 5 | 38602871 | 6,488780 |
| 4766913 | 8 | 146969793 | 6,338480 |
| 4777807 | 5 | 6306024 | 5,744462 |
| 2468079 | 9 | 103283570 | |
| 4777585 | 2 | 12636794 | |
| 4580437 | 5 | 86234114 | |
| 4768105 | 4 | 514726 | |
| 4771134 | 8 | 101470047 | |
| 4776930 | 11 | 48723996 | |

Fonte: Da autora (2019)

Para comparar os métodos modificados com a GWAS, na Figura 4.34, estão identificados os valores de efeito dos marcadores, pelo do gráfico *Manhattan plot*. O eixo Y representa o $-\log_{10}(\text{valor} - p)$ ou $\text{valor} - p$ do teste de Wald. O eixo X é representado pelo SNP, dentro de seu respectivo cromossomo. Em cada cromossomo, os SNPs são ordenados por marca e posição. Foram estabelecidas linhas de corte tracejadas, para identificar os SNPs selecionados, usando valores $-\log_{10}(0,05/23153) = 5,665637$ e $\text{valor} - p = 3,84$ para o teste de Wald.

As análises com o método GWAS, comparadas com os métodos RR-BLUP modificado, Bayes A modificado, Bayes B modificado e BSSV modificado, para as análises de associação genômica, não atingiram o limiar de significância pela correção de Bonferroni e teste de Wald na análise conjunta.

Analisado o método FASTmrMLM em comparação com os métodos RR-BLUP modificado, Bayes A modificado, Bayes B modificado e BSSV modificado, o número de SNPs, significativamente associado à característica de resistência à podridão da espiga, foi cinco e detectado pelo método FASTmrMLM, apresentado na Tabela 4.14.

Contudo a avaliação dos sintomas da doença, relacionando a associação e a seleção de linhagem, constitui um grande desafio para os produtores. A dificuldade de obtenção de métodos confiáveis dificulta ainda mais a associação e seleção de genótipos superiores. Pode ter contribuído, para a redução da capacidade de identificar marcadores significativos pelos métodos modificados, a confirmação da doença com baixa herdabilidade $h^2 = 0,056$.

A análise comparativa com o método FASTmrEMMA não foi possível ser concluída devido a não convergência das cadeias, com uso do pacote mrMLM.

4.3 Considerações Finais

Como mostrado em nosso estudo, a escolha do método que melhor descreve a arquitetura genética pode ser decisiva ao adotar uma estratégia de melhoramento. De forma geral, apesar de os métodos de GWS terem poderes preditivos similares, os padrões de contração dos métodos são muito variáveis, o que permite que informações genéticas sejam recuperadas. O estudo de simulação mostrou que os métodos propostos, com modificações, a partir da marginalização dos efeitos, são eficientes em encontrar genes de interesse em populações estruturadas, ou seja, os métodos são válidos para a capacidade de identificar genes candidatos. Por exemplo, ficou evidente que os métodos Bayes A modificado, Bayes B modificado e BSSV modificado

capturaram informações sobre relações genéticas mais eficientemente, além de recuperar informações sobre o desequilíbrio de ligação de marcadores.

Também ficou notório que o método GWAS apresentou uma alta taxa de falsos positivos em comparação com os métodos propostos modificados (GWSAS), o que leva a uma desvantagem ao uso da GWAS. Em razão da alta capacidade de prever e identificar genes de interesse genômico, o que foi confirmado pelos resultados de simulação, sugere-se que os métodos propostos modificados GWSAS sejam aplicados a estudos de seleção e associação.

O modelo, representado pelo método RR-BLUP e RR-BLUP modificado, produziu um forte encolhimento nos efeitos dos marcadores, à por suposição de variância comum, que pode ter contribuído para a sua reduzida capacidade de identificar genes de interesse.

Os marcadores com maior capacidade de identificação e efeitos significativos podem ser obtidos pela adoção do teste de Wald juntamente com os métodos propostos .

Dentre os métodos propostos para confirmar as hipótese do trabalho, o método Bayes $C\pi$ foi utilizado e submetido a análises de dados simulados, e um padrão equiparável foi observado entre os métodos Bayes $C\pi$ e Bayes $C\pi$ modificado, de forma que estes métodos não foram publicados.

Os resultados com dados reais de milho inspiraram novas análises. No caso, os dados do experimento do milho, avaliados 230 linhagens, foram submetidos a análises usando os 23153 marcadores SNPs genotipados e simulados os fenótipos, considerando herdabilidade 0,45. Esse estudo com interesse comparativo está sendo implementado e terá seus resultados correspondentes apresentados em trabalhos futuros.

Todos os programas de simulação, resultados analíticos, tabelas e gráficos complementares estão disponíveis no apêndice em anexo.

5 CONCLUSÃO

Nosso estudo abordou a questão de unificar as metodologias GWS e GWAS, examinadas como os efeitos do marcadores podem ser estimados pela distribuição a posteriori condicional completa marginalizada. Mostrou-se, analiticamente, que a matriz de covariância genômica entre os genótipos leva em consideração a estrutura da população, e o marcador a ser ajustado não compartilha informação do efeito poligênico, ou seja, a marginalização reduz um número de parâmetros desnecessários no modelo e, assim, torna o ajuste do modelo viável e mais tratável.

Nossos resultados mostraram que os métodos modificados GWSAS, a partir dos métodos clássicos de GWS, podem ser aplicados como método de avaliação na seleção genômica e associação genômica, para a identificação de genes candidatos associados a características de interesse e predição, fornecendo melhores estimativas dos efeitos dos SNPS e predição dos valores fenotípicos e genéticos.

A proposta de testes regularizador de Wald para os efeitos aleatórios, pode ser adotada afim de melhorar a capacidade de identificação de efeitos significativos.

REFERÊNCIAS

- BENJAMINI, Y.; HOCHBERG, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. **Journal of the Royal Statistical Society. Series B. Methodological**, London, v. 57, n. 1, p. 289–300, 1995.
- BERNARDO, J. M.; SMITH, A. F. **Bayesian Theory**. New York: John Wiley & Sons, 1994. 610 p.
- BESAG, J. Spatial interaction and the statistical analysis of lattice systems. **Journal of the Royal Statistical Society. Series B. Methodological**, london, v. 36, p. 192–236, 1974.
- BUSH, W. S.; MOORE, J. H. Genome-wide association studies. **PLoS computational Biology**, San Francisco, v. 8, n. 12, p. e1002822, 2012.
- CANTOR, R. M.; LANGE, K.; SINSHEIMER, J. S. Prioritizing gwas results: a review of statistical methods and recommendations for their application. **American Journal of Human Genetics**, Baltimore, v. 86, n. 1, p. 6–22, Jan 2010.
- CASELLA, G.; GEORGE, E. I. Explaining the gibbs sampler. **The American Statistician**, Washington, v. 46, n. 3, p. 167–174, Aug 1992.
- CHIB, S.; GREENBERG, E. Understanding the metropolis-hastings algorithm. **The American Statistician**, Washington, v. 49, n. 4, p. 327–335, Nov 1995.
- CRUZ, C. D. Genes - a software package for analysis in experimental statistics and quantitative genetics. **Acta Scientiarum**, Maringá, v. 38, n. 3, Jul/Set 2013.
- DAETWYLER, H. D. et al. The impact of genetic architecture on genome-wide evaluation methods. **Genetics**, v. 185, n. 3, p. 1021–1031, July 2010.
- DE LOS CAMPOS, G. et al. Semi-parametric genomic-enabled prediction of genetic values using reproducing kernel hilbert spaces methods. **Genet Research**, Cambridge, v. 92, n. 4, p. 295–308, Aug 2010.
- DE LOS CAMPOS, G. et al. Whole-genome regression and prediction methods applied to plant and animal breeding. **Genetics**, Austin, v. 193, n. 2, p. 327–345, 2013.
- DE LOS CAMPOS, G. et al. Predicting quantitative traits with regression models for dense molecular markers and pedigree. **Genetics**, Austin, v. 182, n. 1, p. 375–385, May 2009.
- DE LOS CAMPOS, G. et al. Prediction of complex human traits using the genomic best linear unbiased predictor. **PLoS Genetics**, San Francisco, v. 9, n. 7, p. e1003608, July 2013.
- DESTA, Z. A.; ORTIZ, R. Genomic selection: genome-wide prediction in plant improvement. **Trends in Plant Science**, Kidlington, v. 19, n. 9, p. 592–601, Sept 2014.
- DOS SANTOS, J. P. R. et al. Genomic selection to resistance to *Stenocarpella maydis* in maize lines using dartseq markers. **BMC Genetics**, London, v. 17, n. 1, p. 86, 2016.
- ENDELMAN, J. B. Ridge regression and other kernels for genomic selection with r package rrblup. **The Plant Genome**, Oxford, v. 4, n. 3, p. 250–255, Nov 2011.

- FRIEDMAN, J.; HASTIE, T.; TIBSHIRANI, R. Regularization paths for generalized linear models via coordinate descent. **Journal of Statistical Software**, Los Angeles, v. 33, n. 1, p. 1–22, 2010.
- GARRICK, D. J.; FERNANDO, R. L. Implementing a qtl detection study (gwas) using genomic prediction methodology. In: **Genome-wide association studies and genomic prediction**. Amsterdam: Springer, 2013. p. 275–298.
- GELMAN, A. et al. **Bayesian data analysis**. Boca Raton: CRC press, 2014. v. 2.
- GEMAN, S.; GEMAN, D. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. **IEEE Transactions on pattern analysis and machine intelligence**, New York, n. 6, p. 721–741, Nov 1984.
- GIANOLA, D. Priors in whole-genome regression: The bayesian alphabet returns. **Genetics**, Austin, v. 194, n. 3, p. 573–596, July 2013.
- GIANOLA, D. et al. Additive genetic variability and the bayesian alphabet. **Genetics**, Austin, v. 183, n. 1, p. 347–363, Sept 2009.
- GIANOLA, D. et al. Genome-wide association studies with a genomic relationship matrix: A case study with wheat and arabidopsis. **G3**, Bethesda, v. 6, n. 10, p. 3241–3256, 2016.
- GIANOLA, D.; FERNANDO, R. L.; STELLA, A. Genomic-assisted prediction of genetic value with semiparametric procedures. **Genetics**, Austin, v. 173, n. 3, p. 1761–1776, July 2006.
- GIANOLA, D.; HOSPITAL, F.; VERRIER, E. Contribution of an additive locus to genetic variance when inheritance is multi-factorial with implications on interpretation of gwas. **Theoretical and Applied Genetics**, Berlin, v. 126, n. 6, p. 1457–1472, June 2013.
- GIANOLA, D.; KAAM, J. B. van. Reproducing kernel hilbert spaces regression methods for genomic assisted prediction of quantitative traits. **Genetics**, Austin, v. 178, n. 4, p. 2289–2303, 2008.
- GIANOLA, D.; ROSA, G. J. One hundred years of statistical developments in animal breeding. **Annual Review of Animal Bioscience**, Palo Alto, v. 3, n. 1, p. 19–56, 2015.
- GOLUB, G. H.; VAN LOAN, C. F. **Matrix computations**. Baltimore: J.H.U. Press, 2012. v. 3.
- GONDRO, C. et al. Quality control for genome-wide association studies. **Methods in Molecular Biology**, Clifton, v. 1019, p. 129–147, 2013.
- HABIER, D. et al. Extension of the bayesian alphabet for genomic selection. **BMC bioinformatics**, London, v. 12, n. 1, p. 186, July 2011.
- HAYES, B.; GODDARD, M. et al. Prediction of total genetic value using genome-wide dense marker maps. **Genetics**, Austin, v. 157, n. 4, p. 1819–1829, Apr 2001.
- HEFFNER, E. L. et al. Genomic selection accuracy for grain quality traits in biparental wheat populations. **Crop Science**, Madison, v. 51, n. 51, p. 2597–2606, Nov 2011.
- HEFFNER, E. L.; JANNINK, J.-L.; SORRELLS, M. E. Genomic selection accuracy using multifamily prediction models in a wheat breeding program. **The Plant Genome**, Oxford, v. 4, n. 1, p. 65–75, Mar 2011.

HENDERSON, C. R. Estimation of variance and covariance components. **Biometrics**, Washington, v. 9, n. 2, p. 226–252, 1953.

HENDERSON, C. R. **Applications of Linear Models in Animal Breeding**. Guelph: University of Guelph, 1984. 462 p.

HOFFMAN, G. E. Correcting for population structure and kinship using the linear mixed model: theory and extensions. **PLoS One**, San Francisco, v. 8, n. 10, p. e75707, 2013.

HOGGART, C. J. et al. Simultaneous analysis of all snps in genome-wide and re-sequencing association studies. **PLoS Genetics**, San Francisco, v. 4, n. 7, p. e1000130, 2008.

HOLLIDAY, J. A.; WANG, T.; AITKEN, S. Predicting adaptive phenotypes from multilocus genotypes in sitka spruce (*picea sitchensis*) using random forest. **G3**, Bethesda, v. 2, n. 9, p. 1085–1093, Sept 2012.

JANNINK, J. L.; LORENZ, A.; IWATA, H. Genomic selection in plant breeding: from theory to practice. **Briefings in Functional Genomics**, Oxford, v. 9, n. 2, p. 166–77, Mar 2010.

KANG, H. M. et al. Variance component model to account for sample structure in genome-wide association studies. **Nature Genetics**, New York, v. 42, n. 4, p. 348–354, 2010.

KANG, H. M. et al. Efficient control of population structure in model organism association mapping. **Genetics**, Austin, v. 178, n. 3, p. 1709–1723, Mar 2008.

KLASEN, J. R. et al. A multi-marker association method for genome-wide association studies without the need for population structure correction. **Nature communications**, New York, v. 7, p. 13299, Nov 2016.

KNIGHT, E. **Improved iterative schemes for REML estimation of variance parameters in linear mixed models**. 298 p. Tese (PhD) — School of Agriculture, Food and Wine, Adelaide, 2008.

LEE, Y.; NELDER, J. A.; PAWITAN, Y. **Generalized linear models with random effects: unified analysis via H-likelihood**. New York: C.R.C. Press, 2006. 416 p.

LEISERSON, M. D. et al. Network analysis of gwas data. **Current Opinion in Genetics & Development**, London, v. 23, n. 6, p. 602–610, Sept 2013.

LI, Z.; SILLANPÄÄ, M. J. Overview of lasso-related penalized regression methods for quantitative trait mapping and genomic selection. **Theoretical and Applied Genetics**, Berlin, v. 125, n. 3, p. 419–35, Aug 2012.

LIPKA, A. E. et al. Gapit: genome association and prediction integrated tool. **Bioinformatics**, Oxford, v. 28, n. 18, p. 2397, Sept 2012.

LIPPERT, C. et al. Fast linear mixed models for genome-wide association studies. **Nature Methods**, New York, v. 8, n. 10, p. 833–835, Apr 2011.

LOH, P.-R. et al. Efficient bayesian mixed-model analysis increases association power in large cohorts. **Nature Genetics**, New York, v. 47, n. 3, p. 284–290, Mar 2015.

- MEUWISSEN, T. H. E.; HAYES, B. J.; GODDARD, M. E. Prediction of total genetic value using genome-wide denbse marker maps. **Genetics**, Austin, v. 157, n. 4, p. 1819–1829, Apr 2001.
- PATTERSON, H. D.; THOMPSON, R. Recovery of inter-block information when block sizes are unequal. **Biometrika**, Oxford, v. 58, n. 3, p. 545–554, Dec 1971.
- PÉREZ, P.; CAMPOS, G. de los. Genome-wide regression and prediction with the bglr statistical package. **Genetics**, Austin, v. 198, n. 2, p. 483–495, Oct 2014.
- POWELL, P. M. V. J. E.; GODDARD, M. E. Reconciling the analysis of ibd and ibs in complex trait studies. **Nature Reviews Genetics**, London, v. 11, n. 11, p. 800–805, Nov 2010.
- PRICE, A. L. et al. Principal components analysis corrects for stratification in genome-wide association studies. **Nature Genetics**, New York, v. 38, n. 8, p. 904, Aug 2006.
- PRICE, A. L. et al. New approaches to population stratification in genome-wide association studies. **Nature Reviews Genetics**, London, v. 11, n. 7, p. 459–463, July 2010.
- R DEVELOPMENT CORE TEAM. **R: A Language and Environment for Statistical Computing**. Vienna, 2016. Disponível em: <<https://www.R-project.org/>>.
- RAKITSCH, B. et al. A lasso multi-marker mixed model for association mapping with population structure correction. **Bioinformatics**, Oxford, v. 29, n. 2, p. 206–214, Jan 2012.
- RSTUDIO. **RStudio: Integrated Development Environment for R**. Boston, 2016. Acesso em 15 mar 2019. Disponível em: <<http://www.rstudio.com/>>.
- RUTKOSKI, J. E.; HEFFNER, E. L.; SORRELLS, M. E. Genomic selection for durable stem rust resistance in wheat. **Wageningen**, Springer, v. 179, n. 1, p. 161–173, May 2011.
- SEGURA, V. et al. An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. **Nature Genetics**, New York, v. 44, n. 7, p. 825, July 2012.
- SHERMAN, J.; MORRISON, W. J. Adjustment of an inverse matrix corresponding to a change in one element of a given matrix. **The Annals of Mathematical Statistics**, San Francisco, v. 21, n. 1, p. 124–127, Mar 1950.
- SINGH, B.; SINGH, A. K. **Marker-assisted plant breeding: principles and practices**. Índia: Springer, 2015. 514 p.
- SLAVOV, G. T. et al. Genome-wide association studies and prediction of 17 traits related to phenology, biomass and cell wall composition in the energy grass miscanthus sinensis. **New Phytologist**, London, v. 201, n. 4, p. 1227–1239, Mar 2014.
- SORENSEN, D.; GIANOLA, D. **Likelihood, Bayesian, and MCMC methods in quantitative genetics**. Amsterdan: Springer Science & Business Media, 2007. 740 p.
- TAMBA, C. L.; NI, Y.-L.; ZHANG, Y.-M. Iterative sure independence screening em-bayesian lasso algorithm for multi-locus genome-wide association studies. **PLoS Computational Biology**, San Francisco, v. 13, n. 1, p. e1005357, 2017.

- TAMBA, C. L.; ZHANG, Y.-M. A fast mrrlm algorithm for multi-locus genome-wide association studies. **bioRxiv**, New York, p. 1–34, June 2018.
- TEMPELMAN, R. J. Statistical and computational challenges in whole genome prediction and genome-wide association analyses for plant and animal breeding. **Journal of Agricultural, Biological, and Environmental Statistics**, Alexandria, v. 20, n. 4, p. 442–466, Dec 2015.
- VANRADEN, P. M. Efficient methods to compute genomic predictions. **Journal of Dairy Science**, Lancaster, v. 91, n. 11, p. 4414–4423, Nov 2008.
- VISSCHER, P. M. et al. 10 years of gwas discovery: biology, function, and translation. **American Journal of Human Genetics**, Baltimore, v. 101, n. 1, p. 5–22, July 2017.
- WANG, C. et al. Ancestry estimation and control of population stratification for sequence-based association studies. **Nature Genetics**, New York, v. 46, n. 4, p. 409, Apr 2014.
- WANG, S.-B. et al. Improving power and accuracy of genome-wide association studies via a multi-locus mixed linear model methodology. **Scientific Reports**, London, v. 6, p. 19444–19444, Apr 2016.
- WEN, Y. et al. Methodological implementation of mixed linear models in multi-locus genome-wide association studies. **Briefings in Bioinformatics**, London, v. 19, p. 700–712, 2017.
- WHITTAKER, J. C.; THOMPSON, R.; DENHAM, M. C. Marker-assisted selection using ridge regression. **Genetics Research**, London, v. 75, n. 2, p. 249–252, Apr 2000.
- XU, S. Estimating polygenic effects using markers of the entire genome. **Genetics**, Austin, v. 163, n. 2, p. 789–801, Feb 2003.
- XU, S. An expectation–maximization algorithm for the lasso estimation of quantitative trait locus effects. **Heredity**, London, v. 105, n. 5, p. 483, Nov 2010.
- XU, S. **Principles of statistical genomics**. Amsterdam: Springer, 2013. 428 p.
- YANG, J. et al. Common snps explain a large proportion of the heritability for human height. **Nature genetics**, New York, v. 42, n. 7, p. 565–569, June 2010.
- YANG, J. et al. Advantages and pitfalls in the application of mixed-model association methods. **Nature Genetics**, New York, v. 46, n. 2, p. 100, Feb 2014.
- YI, N.; GEORGE, V.; ALLISON, D. B. Stochastic search variable selection for identifying multiple quantitative trait loci. **Genetics**, Austin, v. 164, n. 3, p. 1129–1138, July 2003.
- YI, N.; XU, S. Bayesian lasso for qtl mapping. **Genetics**, Austin, June 2008.
- YU, J.; BUCKLER, E. S. Genetic association mapping and genome organization of maize. **Current Opinion in Biotechnology**, London, v. 17, n. 2, p. 155–160, Apr 2006.
- YU, J. et al. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. **Nature Genetics**, New York, v. 38, n. 2, p. 203–208, Feb 2006.
- ZHANG, Z. et al. Mixed linear model approach adapted for genome-wide association studies. **Nature Genetics**, New York, v. 42, n. 4, p. 355–360, 2010.

ZHOU, X.; CARBONETTO, P.; STEPHENS, M. Polygenic modeling with bayesian sparse linear mixed models. **PLOS Genetics**, San Francisco, v. 9, n. 2, p. 1–14, Feb 2013.

ZHOU, X.; STEPHENS, M. Genome-wide efficient mixed-model analysis for association studies. **Nature Genetics**, New York, v. 44, n. 2, p. 821–824, June 2012.

ZHU, C. et al. Status and prospects of association mapping in plants. **The plant genome**, Oxford, v. 1, n. 1, p. 5–20, July 2008.

ZOU, H.; HASTIE, T. Regularization and variable selection via the elastic net. **Journal of the Royal Statistical Society: Series B. Statistical Methodology**, Oxford, v. 67, n. 2, p. 301–320, 2005.

APÊNDICE A – MODELO MÚLTIPLAS MARCAS PARA GWAS

Assume-se que

$$\mathbf{y} \sim N(\mathbf{X}\boldsymbol{\beta} + \mathbf{z}_1 a_1 + \dots + \mathbf{z}_k a_k, \mathbf{I}\sigma_e^2)$$

Também admite-se as seguintes distribuições a priori para cada parâmetro do modelo:

$$\begin{aligned} \boldsymbol{\beta} &\sim N(0, \sigma_\beta^2) \\ a_i &\sim N(0, \sigma_{a_i}^2) \quad i = 1, \dots, k \end{aligned}$$

No segundo estágio da hierarquia, as funções distribuição qui-quadrado invertida escalonada a priori para os hiperparâmetros:

$$\sigma_{a_i}^2 | v_i, s_i^2 \sim \chi^{-2}(v_i = 1, s_i^2 = 0) \quad i = 1, \dots, k$$

Para cada uma das k distribuições, consideramos os valores de $v_i = 1$ e $s_i^2 = 0$. Perceba que se $v_i = 2$ e $s_i^2 = 0$ a distribuição reduz-se a uma distribuição uniforme imprópria. Considerando a condicional posteriori para os marcadores e $\mathbf{R} = \mathbf{I}\sigma_e^2$, usa-se:

$$p(\mathbf{a} | \boldsymbol{\beta}, \sigma_a^2, \sigma_e^2, \mathbf{y}) \propto e^{-\frac{1}{2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{z}_1 a_1 - \dots - \mathbf{z}_k a_k)' \mathbf{R}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{z}_1 a_1 - \dots - \mathbf{z}_k a_k)} p(a_1) \dots p(a_k)$$

denotando por $\tilde{\mathbf{y}} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}$, e na sequência note que escrevemos $[\mathbf{z}'_k \mathbf{R}^{-1} \mathbf{z}_k + (\sigma_{a_k}^2)^{-1}] = \Omega^{-1}$, então para o i -ésimo SNP como um parâmetro de interesse ajustado aos $k-1$ outros parâmetros perturbadores considera-se:

$$\begin{aligned} p(a_i | \boldsymbol{\beta}, a_{-i}, \sigma_a^2, \sigma_e^2, \mathbf{y}) &\propto C \int e^{-\frac{\left[\tilde{\mathbf{y}} - \sum_{i=1}^{k-1} \mathbf{z}_i a_i - \mathbf{z}_k a_k \right]' \mathbf{R}^{-1} \left[\tilde{\mathbf{y}} - \sum_{i=1}^{k-1} \mathbf{z}_i a_i - \mathbf{z}_k a_k \right]}{2}} (\sigma_{a_k}^2)^{-\frac{1}{2}} e^{-\frac{a_k (\sigma_{a_k}^2)^{-1} a_k}{2}} da_k \\ &\propto C e^{-\frac{\left[\tilde{\mathbf{y}} - \sum_{i=1}^{k-1} \mathbf{z}_i a_i \right]' \left[\mathbf{R}^{-1} + \mathbf{R}^{-1} \mathbf{z}_k (\mathbf{z}'_k \mathbf{R}^{-1} \mathbf{z}_k + (\sigma_{a_k}^2)^{-1})^{-1} \mathbf{z}'_k \mathbf{R}^{-1} \right] \left[\tilde{\mathbf{y}} - \sum_{i=1}^{k-1} \mathbf{z}_i a_i \right]}{2}} \\ &\propto C e^{-\frac{\left[\tilde{\mathbf{y}} - \sum_{i=1}^{k-1} \mathbf{z}_i a_i \right]' \left[\mathbf{z}_k \mathbf{z}'_k \sigma_{a_k}^2 + \mathbf{R} \right]^{-1} \left[\tilde{\mathbf{y}} - \sum_{i=1}^{k-1} \mathbf{z}_i a_i \right]}{2}} \end{aligned}$$

Em que $C = \prod_{j=1}^{k-1} (\sigma_{a_j}^2)^{-\frac{1}{2}} e^{-\frac{a_j(\sigma_{a_j}^2)^{-1}a_j}{2}}$. A marginal em relação aos outros $k-1$ SNPs, a ser empregada é dado por:

$$\begin{aligned}
 p(a_i | \boldsymbol{\beta}, \mathbf{a}_{-i}, \sigma_{a_{k-1}}^2, \sigma_e^2) &\propto (\sigma_{a_i}^2)^{-\frac{1}{2}} e^{-\frac{a_i(\sigma_{a_i}^2)^{-1}a_i}{2}} \iint \dots \int e^{-\frac{\left[\bar{y} - \sum_{j=1}^{k-1} \mathbf{z}_j a_j - \mathbf{z}_i a_i \right]' \mathbf{R}^{-1} \left[\bar{y} - \sum_{j=1}^{k-1} \mathbf{z}_j a_j - \mathbf{z}_i a_i \right]}{2}} \prod_{j=1}^{k-1} (\sigma_{a_j}^2)^{-\frac{1}{2}} e^{-\frac{a_j(\sigma_{a_j}^2)^{-1}a_j}{2}} da_j \\
 &\propto (\sigma_{a_i}^2)^{-\frac{1}{2}} e^{-\frac{a_i(\sigma_{a_i}^2)^{-1}a_i}{2}} e^{-\frac{(\bar{y} - \mathbf{z}_i a_i)' \left[\sum_{j=1}^{k-1} \mathbf{z}_j \mathbf{z}_j' \sigma_{a_j}^2 + \mathbf{R} \right]^{-1} (\bar{y} - \mathbf{z}_i a_i)}{2}} \\
 &\propto (\sigma_{a_i}^2)^{-\frac{1}{2}} e^{-\frac{a_i(\sigma_{a_i}^2)^{-1}a_i}{2}} e^{-\frac{(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{z}_i a_i)' [\mathbf{z}_{k-1} \boldsymbol{\Psi} \mathbf{z}_{k-1}' + \mathbf{R}]^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{z}_i a_i)}{2}}
 \end{aligned}$$

no qual \mathbf{Z}_{k-1} é a matriz de SNPs $n \times (k-1)$ e a informação do i -ésimo marcador é ausente. A matriz diagonal assume variância específica para cada marcador.

$$a_i | \boldsymbol{\beta}, a_{-i}, \sigma_{a_{k-1}}^2, \sigma_e^2, \mathbf{y} \sim N \left[(\mathbf{z}_i' \mathbf{V}_{k-1}^{-1} \mathbf{z}_i + (\sigma_{a_i}^2)^{-1})^{-1} \mathbf{z}_i' \mathbf{V}_{k-1}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}), [\mathbf{z}_i' \mathbf{V}_{k-1}^{-1} \mathbf{z}_i + (\sigma_{a_i}^2)^{-1}]^{-1} \right] \quad (1)$$

ANEXO A - CÓDIGOS UTILIZADOS NOS PROCEDIMENTOS DE SIMULAÇÃO E ANÁLISE PELO SOFTWARE R

```
#####
#Simular fenótipos
# Código R reproduzível com propósito ilustrativo de simulação e análise de dados
#####
# arquivos
myGD = read.table("GD.txt",head = TRUE) # arquivo matriz de marcadores genotipados
M     = myGD[,2:ncol(myGD)]
M     = as.matrix(M)
ncol(M) #número de SNPs

# se cenário oligogênico (0,1% dos genes), descomente e execute linha abaixo:
#ngen = ncol(M)*0.001
# ou
# se cenário poligênico (1% dos genes), descomente e execute linha seguinte:
#ngen = col(M)*0.01

genes      = sample(1:ncol(M),ngen,replace=F) # genes amostrados
efe        = matrix(0,ncol(M),1)           # vetor reter os efeito dos genes
efe[genes] = rnorm(ngen)                   # efeitos simulados de uma distribuição normal padrão
dim(efe)   =                               # tamanho do vetor de efeitos
efe        = as.matrix(efe)

# centralizando matriz de genótipos
freq       = (colMeans(M))/2                # média para cada marcador
Fa         = t(matrix(freq,ncol(M), nrow(M))) # média frequência alélica
geno       = M-2*Fa                         # matriz de marcadores corrigida pela média
vvg        = geno%*%efe                     # valor genético genômico
mean(vvg)  =                               # média do VGG
vg         = var(vvg)*(nrow(efe)-1)/nrow(efe) # estimando a máxima variância pelo estimador

#Herdabilidades consideradas: h=0.2, h=0.5 e h=0.8;
ve         = (vg-h*vg)/h # Substituir h pelo valor desejado e executar linha
y          = vgg+rnorm(nrow(M))*sqrt(ve)     #cálculo dos valores observados

#salvando arquivos
write.table(genes,"genes.txt")              # salvar arquivo
write.table(efe,"efeito.txt")              # salvar arquivo
write.table(vvg,"vvg.txt")                 # salvar arquivo
write.table(mean(vvg),"mean(vvg).txt")     # salvar arquivo
write.table(vg,"vg.txt")                   # salvar arquivo
write.table(ve,"veo2.txt")                 # salvar arquivo
write.table(y,"yo2.txt")                   # salvar arquivo fenótipo simulado

#####
# RR-BLUP modificado
```

```

# Código R reproduzível com propósito ilustrativo de simulação e análise de dados
#####
# arquivos
myGD = read.table("GD.txt",head = TRUE) # arquivo matriz de marcadores genotipados
M     = myGD[,2:ncol(myGD)]
M     = as.matrix(M)
myY   = read.table("yo2.txt",head = TRUE)# vetor de observações fenotipados
Y     = myY[,2:ncol(myY)]
Z     = M

# centralizando matriz de genótipos
freq  = (colMeans(Z))/2           # média para cada marcador
Fa    = t(matrix(freq,ncol(Z), nrow(Z))) # média frequência alélica
W     = Z-2*Fa                   # matriz de marcadores corrigida pela média
Z     = W                         # renomeando a matriz W
y     = as.matrix(y)             # y matricial
n     = nrow(y)                  # número de observações
x     = matrix(1,n,1)            # matriz de efeito fixo

# parâmetros
var   = var(y)                   # valor inicial variância fenotípica
beta  = solve(t(x)%*%x)%*%t(x)%*%y # valor inicial para efeito fixo
niter = 3000                      # número de iterações
bu    = 300                       # iterações iniciais descartadas
v     = 1                          # hiperparâmetro (grau de liberdade) p/variância SNPs
s2    = var(y)*0.005/(ncol(Z))    # hiperparâmetro (escala) p/ variância SNPs
salto = seq(bu,niter,3)           # salto definido pelas iterações

# matrizes repositórios
a     = matrix(0,ncol(Z),1)       # matriz reter efeitos aleatórios amostrados
wald  = matrix(0,1,ncol(Z))      # matriz reter valores teste wald
wald2 = matrix(0,1,ncol(Z))      # matriz reter valores teste wald
va    = matrix(var/ncol(Z),1,ncol(Z)) # matriz reter variância genética amostrada com valor inicial var/ncol(z)
ipslom = s2/ncol(Z)              # variância para marcador individual 5% variancia fenotípica
am    = a                         # matriz contém o valor efeito aleatório

# amostrando MCMC
for(iter in 1:niter)
{
G = Z*matrix(rep(va,n),n,byrow=T) # matriz de parentesco ajustada a partir da amostra da variância genética
V = tcrossprod(G,Z)+diag(1,n)*c(var) # matriz de covariância de y, forma V=ZGZ'+R
Vi = solve(V)                      # inversão matriz V
beta1 = solve(t(x)%*%Vi%*%x)%*%t(x)%*%Vi%*%y) # amostrando efeito fixo (média)
vb    = solve(t(x)%*%Vi%*%x)        # inversão X'ViX (variância efeito fixo)
beta  = beta1+chol(vb)%*%rnorm(nrow(beta)) # acumular soma p/ calculo media posteriori
yc    = y-x%*%beta                  # ajustando y a partir da amostra de beta

for(i in 1:ncol(Z))                # contador para o número de marcadores ajustados na iteração
{

```

```

I = t(Z[,i])%*%Vi%*%Z[,i] # fórmula z'Viz procedendo transformação Vi
O = t(Z[,i])%*%Vi%*%(yc) # formula z'Viyc procedendo transformação de Vi
PS = 1/(I-(1/va[i])) # matriz escalar de valor inicial
val = solve(I-I*PS*I + (1/va[i])) # variância do marcador
a1 = solve(I-I*PS*I + (1/va[i]))%*(O-I*PS*O) # média do marcador

a2 = sqrt(val)*rnorm(1) +a1 # soma para média marcador posteriori
va[i] = va2 # retendo a variância amostrada para cada marcador
a[i] = a2 # retendo efeito marcador amostrado para cada marcador
wald[,i] = ((a2**2)/val) # retendo cálculo teste Wald para cada marcador
wald2[,i] = (val) # retendo as variâncias do marcador por uma nova nome
}

ab = rchisq(1,v+ncol(Z)) # valor amostrado qui-quadrado(1,v+ncol(Z))
sa = (v*s2+ sum(a^2))/ab # variância genética amostrada de qui-quadrado
ae = rchisq(1,n+v) # valor amostrado qui-quadrado(1,n+v)
theta = x%*%beta+Z%*%a # fenótipo ajustado para cada marcador
var = (v*s2+t(y-theta)%*(y-theta))/ae # cálculo da variância residual
va = matrix(sa,1,ncol(Z)) # armazenamento variância genética de cada marcador
write(iter, "iter1") # arquivando o contador de iterações
am = ( iter*am+a)/(iter +1) #

if ((iter>bu) & (sum(+iter==salto)==1 )){ # sentença lógica encerra o processo MCMC
# arquivando as cadeias
write(t(a),"a2.txt", sep = ";", append=TRUE, ncol=length(a))
write(beta,"beta2.txt", sep = ";", append=TRUE, ncol=length(beta))
write(var,"var.txt", sep = ";", append=TRUE,ncol=length(var))
write(wald,"wald.txt", sep = ";", append=TRUE,ncol=length(wald))
write(wald2,"vadit.txt", sep = ";", append=TRUE,ncol=length(wald))
write(va[1],"va.txt", sep = ";",append=TRUE, ncol=length(va))

}

}

#####
# Bayes A modificado
# Código R reproduzível com propósito ilustrativo de simulação e análise de dados
#####
# arquivos
myGD = read.table("GD.txt",head = TRUE) # arquivo matriz de marcadores genotipados
M = myGD[,2:ncol(myGD)]
M = as.matrix(M)
myY = read.table("yo2.txt",head = TRUE)# vetor de observações fenotipados
Y = myY[,2:ncol(myY)]
Z = M

# centralizando matriz de genótipos
freq = (colMeans(Z))/2 # média para cada marcador
Fa = t(matrix(freq,ncol(Z), nrow(Z))) # média frequência alélica

```

```

W = Z-2*Fa # matriz de marcadores corrigida pela média
Z = W # renomeando a matriz W
y = as.matrix(y) # y matricial
n = nrow(y) # número de observações
x = matrix(1,n,1) # matriz de efeito fixo

# parâmetros
var = var(y) # valor inicial variância fenotípica
beta = solve(t(x)%*%x)%*%t(x)%*%y # valor inicial para efeito fixo
niter = 3000 # número de iterações
bu = 300 # iterações iniciais descartadas
v = 1 # hiperparâmetro (grau de liberdade) p/variância SNPs
s2 = var(y)*0.005/(ncol(Z)) # hiperparâmetro (escala) p/ variância SNPs
salto = seq(bu,niter,3) # salto definido pelas iterações

# matrizes repositórios
a = matrix(0,ncol(Z),1) # matriz reter efeitos aleatórios amostrados
wald = matrix(0,1,ncol(Z)) # matriz reter valores teste wald
wald2 = matrix(0,1,ncol(Z)) # matriz reter valores teste wald
va = matrix(var/ncol(Z),1,ncol(Z)) # matriz reter variância genética amostrada com valor inicial var/ncol(z)

# amostrando MCMC
for(iter in 1:niter)
{
G = Z*matrix(rep(va,n),n,byrow=T) # matriz de parentesco ajustada a partir da amostra da variância genética
V = tcrossprod(G,Z)+diag(1,n)*c(var) # matriz de covariância de y, forma V=ZGZ'+R
Vi = solve(V) # inversão matriz V
beta1 = solve(t(x)%*%Vi%*%x)%*%t(x)%*%Vi%*%y # amostrando efeito fixo (média)
vb = solve(t(x)%*%Vi%*%x) # inversão X'ViX (variância efeito fixo)
beta = beta1+chol(vb)%*%rnorm(nrow(beta)) # acumular soma p/ calculo media posteriori
yc = y-x%*%beta # ajustando y a partir da amostra de beta

for(i in 1:ncol(Z)) # contador para o número de marcadores ajustados na iteração
{
I = t(Z[,i])%*%Vi%*%Z[,i] # fórmula z'Viz procedendo transformação Vi
O = t(Z[,i])%*%Vi%*%(yc) # formula z'Viyc procedendo transformação de Vi
PS = 1/(I-(1/va[i])) # matriz escalar de valor inicial
val = solve(I-I*PS*I + (1/va[i])) # variância do marcador
a1 = val%*%(O-I*PS*O) # média do marcador

a2 = sqrt(val)*rnorm(1) +a1 # soma para média marcador posteriori
ab = rchisq(1,v+1) # valor amostrado qui-quadrado(1,v+1)
va2 = (v*s2+ a2^2)/ab # variância genética amostrada de qui-quadrado
PS2 = ((va2-va[i])/(I*(va2-va[i])+1))
va[i] = va2 # retendo a variância amostrada para cada marcador
a[i] = a2 # retendo efeito marcador amostrado para cada marcador

wald[,i] = ((a2**2)/val) # retendo cálculo teste Wald para cada marcador

```

```

wald2[,i] = (val) # retendo as variâncias do marcador por uma nova nome
Vi = Vi-Vi%*%Z[,i]%*%as.matrix(PS2)%*%t(Z[,i])%*%Vi # Soma da matriz inversa Teorema Sherman-Morrison

}

ae = rchisq(1,n+v) # valor amostrado qui-quadrado(1,v+1)
theta = x%*%beta+Z%*%a # fenótipo ajustado para cada marcador
var = (v*s2+t(y-theta)%*%(y-theta))/ae # cálculo da variância residual
write(iter,"iter1") # arquivando o contador de iterações

if ((iter>bu) & ( sum+(iter==salto)==1) ) { # sentença lógica encerra o processo MCMC

# arquivando as cadeias
write(t(a), "a2.txt", sep=";", append=TRUE, ncol=length(a))
write(beta, "beta2.txt", sep=";", append=TRUE, ncol=length(beta))
write(var, "var.txt", sep=";", append=TRUE, ncol=length(var))
write(wald, "wald.txt", sep=";", append=TRUE, ncol=length(wald))
write(wald2, "vadit.txt", sep=";", append=TRUE, ncol=length(wald))
write(va, "va.txt", sep=";", append=TRUE, ncol=length(va))
}
}

#####
# Bayes B modificado
# Código R reproduzível com propósito ilustrativo de simulação e análise de dados
#####
# arquivos
myGD = read.table("GD.txt", head = TRUE) # arquivo matriz de marcadores genotipados
M = myGD[,2:ncol(myGD)]
M = as.matrix(M)
myY = read.table("yo2.txt", head = TRUE) # vetor de observações fenotipados
y = myY[,2:ncol(myY)]
Z = M

# centralizando matriz de genótipos
freq = (colMeans(Z))/2 # média para cada marcador
Fa = t(matrix(freq, ncol(Z), nrow(Z))) # média frequência alélica
W = Z-2*Fa # matriz de marcadores corrigida pela média
Z = W # renomeando a matriz W
y = as.matrix(y) # y matricial
n = nrow(y) # número de observações
x = matrix(1,n,1) # matriz de efeito fixo

# parâmetros
var = var(y) # valor inicial variância fenotípica
beta = solve(t(x)%*%x)%*%t(x)%*%y # valor inicial para efeito fixo
niter = 3000 # número de iterações
bu = 300 # iterações iniciais descartadas
v = 1 # hiperparâmetro (grau de liberdade) p/variância SNPs

```

```

pi      = 0.5                                # parâmetro "pi" define a probabilidade do efeito do SNP
s2      = var(y)*0.005/(ncol(Z)*(1-pi)) # hiperparâmetro (escala) p/ variância SNPs
salto   = seq(bu,niter,3)                    # salto definido pelas iterações
ipslom  = s2/ncol(Z)                          # variância para marcador individual 5% variancia fenotípica

# matrizes repositórios
a       = matrix(0,ncol(Z),1)                 # matriz reter efeitos aleatórios amostrados
wald    = matrix(0,1,ncol(Z))                # matriz reter valores teste wald
wald2   = matrix(0,1,ncol(Z))                # matriz reter valores
va      = matrix(var/ncol(Z),1,ncol(Z)) # matriz reter variância genética amostrada com valor inicial var/ncol(z)

# amostrando MCMC
for(iter in 1:niter)
{
G = Z*matrix(rep(va,n),n,byrow=T) # matriz de parentesco ajustada a partir da amostra da variância genética
V = tcrossprod(G,Z)+diag(1,n)*c(var) # matriz de covariância de y, forma V=ZGZ'+R
Vi = solve(V) # inversão matriz V
beta1 = solve(t(x)%*%Vi%*%x)%*%t(x)%*%Vi%*%y) # amostrando efeito fixo (média)
vb     = solve(t(x)%*%Vi%*%x) # inversão X'ViX (variância efeito fixo)
beta   = beta1+chol(vb)%*%rnorm(nrow(beta)) # acumular soma p/ calculo media posteriori
yc     = y-x%*%beta # ajustando y a partir da amostra de beta

for(i in 1:ncol(Z)) # contador para o número de marcadores ajustados na iteração
{
I = t(Z[,i])%*%Vi%*%Z[,i] # fórmula z'Viz procedendo transformação Vi
O = t(Z[,i])%*%Vi%*%yc # formula z'Viyc procedendo transformação de Vi
PS = 1/(I-(1/va[i])) # matriz escalar de valor inicial
PSS = 1/(I-(1/ipslom)) # matriz escalar de valor inicial
va1 = solve(I-I*PS*I + (1/va[i])) # variância do marcador
a1 = solve(I-I*PS*I + (1/va[i]))%*%(O-I*PS*O) # média do marcador
a3 = solve(I-I*PSS*I + (1/ipslom))%*%(O-I*PSS*O)
va3 = solve(I-I*PSS*I + (1/ipslom))

a2 = sqrt(va1)*rnorm(1) +a1 # soma para média marcador posteriori
a4 = sqrt(va3)*rnorm(1) +a3
ab = rchisq(1,v+1) # valor amostrado qui-quadrado(1,v+1)

PS2 = ((va2-va[i])/(I*(va2-va[i])+1))
va[i] = va2 # retendo a variância amostrada para cada marcador
a[i] = a2 # retendo efeito marcador amostrado para cada marcador
Va = Vi-Vi%*%Z[,i]%*%as.matrix(PS)%*%t(Z[,i])%*%Vi
SS1 = t(yc-Z[,i]*a2)%*%Va%*%(yc-Z[,i]*a2)/2 # cálculo ai pela normal
ss2 = t(yc-Z[,i]*a4)%*%Va%*%(yc-Z[,i]*a4)/2 # cálculo aj pela normal
p1 = (1-pi)*exp(-(SS1)) # cálculo parcial da distribuição deltai
p2 = pi*exp(-(ss2)) # cálculo parcial da distribuição deltai

if (is.na(p1) | is.na(p2)) {delta=0} else {delta=p1/(p1+p2)}
rest = delta>runif(1)
va2 = (v*s2+ a2^2)/ab # variância genética amostrada de qui-quadrado

```

```

PS2 = (+rest)*((va2-va[i])/(I*(va2-va[i])+1)) + (1-(+rest))*((ipslom-va[i])/(I*(ipslom-va[i])+1))

va[i] = (+rest)*va2 + (1-(+rest))*ipslom
a[i] = (+rest)*a2 + (1-(+rest))*a4
wald[,i] = (+rest)*((a2**2)/val) + (1-(+rest))*(a4**2/va3) # retendo cálculo teste Wald para cada marcador
wald2[,i] = (+rest)*(val) + (1-(+rest))*(va3) # retendo as variâncias do marcador por uma nova nome
Vi = Vi-Vi**Z[,i]**as.matrix(PS2)**t(Z[,i])**Vi # Soma da matriz inversa Teorema Sherman-Morrison

}

ae = rchisq(1,n+v) # valor amostrado qui-quadrado(1,v+1)
theta = x**beta+Z**a # fenótipo ajustado para cada marcador
var = (v*s2+t*(y-theta)**(y-theta))/ae # cálculo da variância residual
write(iter,"iter1") # arquivando o contador de iterações

if ((iter>bu) & (sum(+iter==salto))==1) ) { # sentença lógica encerra o processo MCMC

# arquivando as cadeias
write(t(a), "a2.txt", sep=";", append=TRUE, ncol=length(a))
write(beta, "beta2.txt", sep=";", append=TRUE, ncol=length(beta))
write(var, "var.txt", sep=";", append=TRUE, ncol=length(var))
write(wald, "wald.txt", sep=";", append=TRUE, ncol=length(wald))
write(wald2, "vadit.txt", sep=";", append=TRUE, ncol=length(wald))
write(va, "va.txt", sep=";", append=TRUE, ncol=length(va))
}
}

#####
# BSSV modificado
# Código R reproduzível com propósito ilustrativo de simulação e análise de dados
#####
# arquivos
myGD = read.table("GD.txt", head = TRUE) # arquivo matriz de marcadores genotipados
M = myGD[,2:ncol(myGD)]
M = as.matrix(M)
myY = read.table("yo2.txt", head = TRUE) # vetor de observações fenotipados
y = myY[,2:ncol(myY)]
Z = M

# centralizando matriz de genótipos
freq = (colMeans(Z))/2 # média para cada marcador
Fa = t(matrix(freq, ncol(Z), nrow(Z))) # média frequência alélica
W = Z-2*Fa # matriz de marcadores corrigida pela média
Z = W # renomeando a matriz W
y = as.matrix(y) # y matricial
n = nrow(y) # número de observações
x = matrix(1,n,1) # matriz de efeito fixo

```

```

# parâmetros
var = var(y) # valor inicial variância fenotípica
beta = solve(t(x)%*%x)%*%t(x)%*%y # valor inicial para efeito fixo
niter = 3000 # número de iterações
bu = 300 # iterações iniciais descartadas
v = 1 # hiperparâmetro (grau de liberdade) p/variância SNPs
pi = 0.5 # parâmetro "pi" define a probabilidade do efeito do SNP
s2 = var(y)*0.005/(ncol(Z)*(1-pi)) # hiperparâmetro (escala) p/ variância SNPs
salto = seq(bu,niter,3) # salto definido pelas iterações
ipslom = s2/ncol(Z) # variância para marcador individual 5% variancia fenotípica

# matrizes repositórios
a = matrix(0,ncol(Z),1) # matriz reter efeitos aleatórios amostrados
wald = matrix(0,1,ncol(Z)) # matriz reter valores teste wald
wald2 = matrix(0,1,ncol(Z)) # matriz reter valores
va = matrix(var/ncol(Z),1,ncol(Z)) # matriz reter variância genética amostrada com valor inicial var/ncol(z)
am = a # matriz contém o valor efeito aleatório
pcad = matrix(0,1,ncol(Z))

# amostrando MCMC
for(iter in 1:niter)
{
G = Z*matrix(rep(va,n),n,byrow=T) # matriz de parentesco ajustada a partir da amostra da variância genética
V = tcrossprod(G,Z)+diag(1,n)*c(var) # matriz de covariância de y, forma V=ZGZ'+R
Vi = solve(V) # inversão matriz V
beta1 = solve(t(x)%*%Vi%*%x)%*%t(x)%*%Vi%*%y # amostrando efeito fixo (média)
vb = solve(t(x)%*%Vi%*%x) # inversão X'ViX (variância efeito fixo)
beta = beta1+chol(vb)%*%rnorm(nrow(beta)) # acumular soma p/ calculo media posteriori
yc = y-x%*%beta # ajustando y a partir da amostra de beta

for(i in 1:ncol(Z)) # contador para o número de marcadores ajustados na iteração
{
I = t(Z[,i])%*%Vi%*%Z[,i] # fórmula z'Viz procedendo transformação Vi
O = t(Z[,i])%*%Vi%*%(yc) # formula z'Viyc procedendo transformação de Vi
PS = 1/(I-(1/va[i])) # matriz escalar de valor inicial
PSS = 1/(I-(1/ipslom)) # matriz escalar de valor inicial
va1 = solve(I-I*PS*I + (1/va[i]) ) # variância do marcador
a1 = solve(I-I*PS*I + (1/va[i]) )%*%(O-I*PS*O) # média do marcador
a3 = solve(I-I*PSS*I + (1/ipslom) )%*%(O-I*PSS*O)
va3 = solve(I-I*PSS*I + (1/ipslom) )

a2 = sqrt(va1)*rnorm(1) +a1 # soma para média marcador posteriori
a4 = sqrt(va3)*rnorm(1) +a3
ab = rchisq(1,v+1) # valor amostrado qui-quadrado(1,v+1)

PS2 = ((va2-va[i])/(I*(va2-va[i])+1))
va[i] = va2 # retendo a variância amostrada para cada marcador
a[i] = a2 # retendo efeito marcador amostrado para cada marcador
Va = Vi-Vi%*%Z[,i]%*%as.matrix(PS)%*%t(Z[,i])%*%Vi

```

```

SS1 = t(yc-Z[,i]*a2)**%Va**% (yc-Z[,i]*a2)/2 # cálculo ai pela normal
ss2 = t(yc-Z[,i]*a4)**%Va**% (yc-Z[,i]*a4)/2 # cálculo aj pela normal
p1 = (1-pi)*exp(-(SS1)) # cálculo parcial da distribuição deltai
p2 = pi*exp(-(ss2)) # cálculo parcial da distribuição deltai

if (is.na(p1) | is.na(p2)) {delta=0} else {delta=p1/(p1+p2)}
rest = delta>runif(1)
va2 = (v*s2+ a2^2)/ab # variância genética amostrada de qui-quadrado
PS2 = (+rest)*((va2-va[i])/(I*(va2-va[i])+1)) + (1-(+rest))*((ipslom-va[i])/(I*(ipslom-va[i])+1))

va[i] = (+rest)*va2 + (1-(+rest))*ipslom
a[i] = (+rest)*a2 + (1-(+rest))*a4
pcad[i] = (+rest)
wald[,i] = (+rest)*((a2**2)/va1) + (1-(+rest))*(a4**2/va3) # retendo cálculo teste Wald para cada marcador
wald2[,i] = (+rest)*(va1) + (1-(+rest))*(va3) # retendo as variâncias do marcador por uma nova nome
Vi = Vi-Vi**%Z[,i]**%as.matrix(PS2)**%t(Z[,i])**%Vi # Soma da matriz inversa Teorema Sherman-Morrison

}

ae = rchisq(1,n+v) # valor amostrado qui-quadrado(1,v+1)
theta = x**%beta+Z**%a # fenótipo ajustado para cada marcador
pies = sum(pcad)
pi = 1-rbeta(1,pies+1,ncol(pcad)-pies+1) # pi amostrado distribuição beta
var = (v*s2+t(y-theta)**%(y-theta))/ae # cálculo da variância residual
write(iter,"iter1") # arquivando o contador de iterações
write(pi, "pi.txt",append=TRUE)
am=( iter*am+a)/(iter +1)

if ((iter>bu) & ( sum+(iter==salto))==1) ) { # sentença lógica encerra o processo MCMC

# arquivando as cadeias
write(t(a), "a2.txt", sep=";", append=TRUE, ncol=length(a))
write(beta, "beta2.txt", sep=";", append=TRUE, ncol=length(beta))
write(var, "var.txt", sep=";", append=TRUE, ncol=length(var))
write(wald, "wald.txt", sep=";", append=TRUE, ncol=length(wald))
write(wald2, "vadt.txt", sep=";", append=TRUE, ncol=length(wald))
write(va, "va.txt", sep=";", append=TRUE, ncol=length(va))
}
}

#####
# BSSV original
# Código R reproduzível com propósito ilustrativo de simulação e análise de dados
#####
# arquivos
myGD = read.table("GD.txt", head = TRUE) # arquivo matriz de marcadores genotipados
M = myGD[,2:ncol(myGD)]
M = as.matrix(M)

```

```

myY = read.table("yo2.txt",head = TRUE)# vetor de observações fenotipados
y   = myY[,2:ncol(myY)]
Z   = M

# centralizando matriz de genótipos
freq = (colMeans(Z))/2           # média para cada marcador
Fa   = t(matrix(freq,ncol(Z), nrow(Z))) # média frequência alélica
W    = Z-2*Fa                   # matriz de marcadores corrigida pela média
Z    = W                        # renomeando a matriz W
y    = as.matrix(y)             # y matricial
n    = nrow(y)                  # número de observações
x    = matrix(1,n,1)            # matriz de efeito fixo

# parâmetros
var   = var(y)                  # valor inicial variância fenotípica
beta  = solve(t(x)%*%x)%*%t(x)%*%y # valor inicial para efeito fixo
niter = 3000                    # número de iterações
bu    = 300                     # iterações iniciais descartadas
v     = 1                       # hiperparâmetro (grau de liberdade) p/variância SNPs
pi    = 0.5                     # parâmetro "pi" define a probabilidade do efeito do SNPs
s2    = var(y)*0.005/(ncol(Z)*(1-pi)) # hiperparâmetro (escala) p/ variância SNPs
salto = seq(bu,niter,3)         # salto definido pelas iterações
ipslom = s2/ncol(Z)            # variância para marcador individual 5% variancia fenotípica

# matrizes repositórios
a     = matrix(0,ncol(Z),1)     # matriz reter efeitos aleatórios amostrados
wald  = matrix(0,1,ncol(Z))    # matriz reter valores teste wald
wald2 = matrix(0,1,ncol(Z))    # matriz reter valores
va    = matrix(var/ncol(Z),1,ncol(Z)) # matriz reter variância genética amostrada com valor inicial var/ncol(z)
am    = a                       # matriz contém o valor efeito aleatório
pcad  = matrix(0,1,ncol(Z))

# amostrando MCMC
for(iter in 1:niter)
{
beta1 = solve(t(x)%*%x)%*%t(x)%*%(y-Z%*%a) # amostrando efeito fixo (média)
vb    = solve(t(x)%*%x)*c(var)              # inversão X'ViX (variância efeito fixo)
beta  = beta1+chol(vb)%*%rnorm(nrow(beta))  # acumular soma p/ calculo media posteriori
yc    = y-x%*%beta                          # ajustando y a partir da amostra de beta

for(i in 1:ncol(Z)) # contador para o número de marcadores ajustados na iteração
{
a1 = solve(t( Z[,j])%*%Z[,j]+c(var/va[j]))%*%t(Z[,j])%*%(y-x%*%beta-Z[,-j]%*%a[-j]) # média do marcador
va1 = sqrt(solve( t( Z[,j])%*%Z[,j]+c(var/va[j]) ) *c(var)) # variância do marcador
a3 = solve(t( Z[,j])%*%Z[,j]+c(var/ipslom))%*%t(Z[,j])%*%(y-x%*%beta-Z[,-j]%*%a[-j])
va3 = solve(I-I*PSS*I + (1/ipslom) )

a2 = va1%*%rnorm(1) +a1 # soma para média marcador posteriori
a4 = va2%*%rnorm(1) +a3

```

```

ab = rchisq(nrow(a2),v) # valor amostrado qui-quadrado (nrow(a2),v)
yc = y-x**beta-Z[,-j]**a[-j]
va2 = sqrt(solve(t( Z[,j]**Z[,j]+c(var/ipslom))*c(var))

SS1 = t(yc- Z[,j]**a2)**(yc- Z[,j]**a2) # cálculo de ai pela normal
ss2 = t(y-Z[,j]**a4)**(y-Z[,j]**a4) # cálculo de aj pela normal
p1 = (1-pi)*exp(-(SS1))/(2*var) # cálculo parcial da distribuição deltai
p2 = pi*exp(-(ss2))/(2*var) # cálculo parcial da distribuição deltai

if (is.na(p1) | is.na(p2)) {delta=0} else {delta=p1/(p1+p2)}
rest = delta>runif(1)
va2 = (v*s2+ a2^2)/ab # variância genética amostrada de qui-quadrado
PS2 = (+rest)*((va2-va[i])/(I*(va2-va[i])+1)) + (1-(+rest))*((ipslom-va[i])/(I*(ipslom-va[i])+1))

va[j] = (v*s2+ (+rest)*a2^2)/ab
a[j] = (+rest)*a2+(1-(+rest))*a4
pcad[i] = (+rest)
wald[,j] = (+rest)*((a2**2)/val) +(1-(+rest))*(a4**2/va2) # retendo cálculo teste Wald para cada marcador
wald2[,j] = (+rest)*(val) +(1-(+rest))*(va2) # retendo as variâncias do marcador por uma nova nome

}

ae = rchisq(1,n+v) # valor amostrado qui-quadrado(1,v+1)
theta = x**beta+Z**a # fenótipo ajustado para cada marcador
pies = sum(pcad)
pi = 1-rbeta(1,pies+1,ncol(pcad)-pies+1) # pi amostrado distribuição beta
var = (v*s2+t(y-theta)**(y-theta))/ae # cálculo da variância residual
write(iter,"iter1") # arquivando o contador de iterações
write(pi, "pi.txt",append=TRUE)

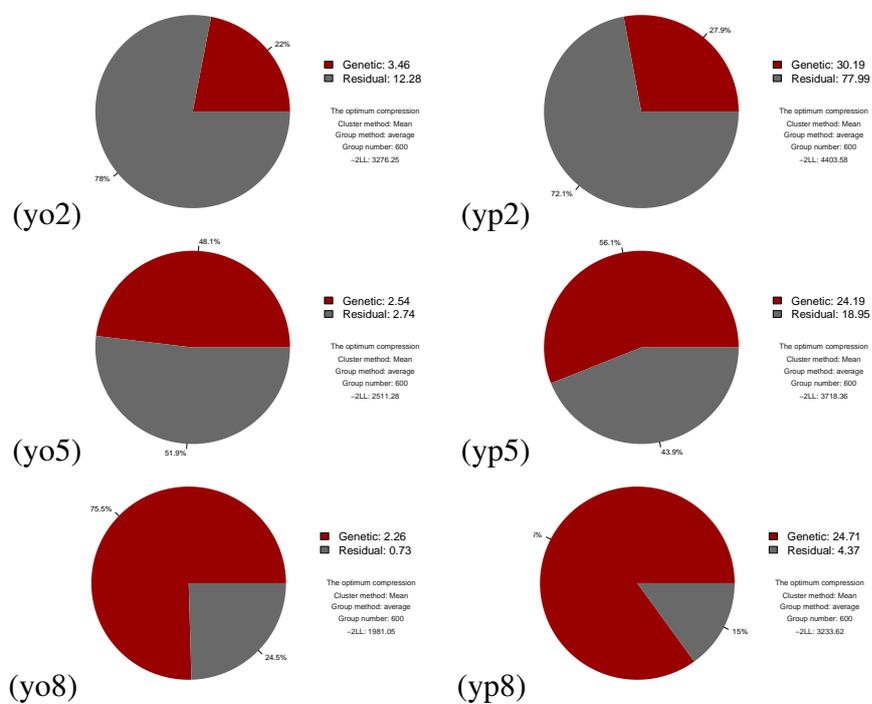
if ((iter>bu) & ( sum(+iter==salto))==1) ){ # sentença lógica encerra o processo MCMC

# arquivando as cadeias
write(t(a), "a2.txt", sep=";", append=TRUE, ncol=length(a))
write(beta, "beta2.txt", sep=";", append=TRUE, ncol=length(beta))
write(var, "var.txt", sep=";", append=TRUE, ncol=length(var))
write(wald, "wald.txt", sep=";", append=TRUE, ncol=length(wald))
write(wald2, "vadit.txt", sep=";", append=TRUE, ncol=length(wald))
write(va, "va.txt", sep=";", append=TRUE, ncol=length(va))
}
}

```

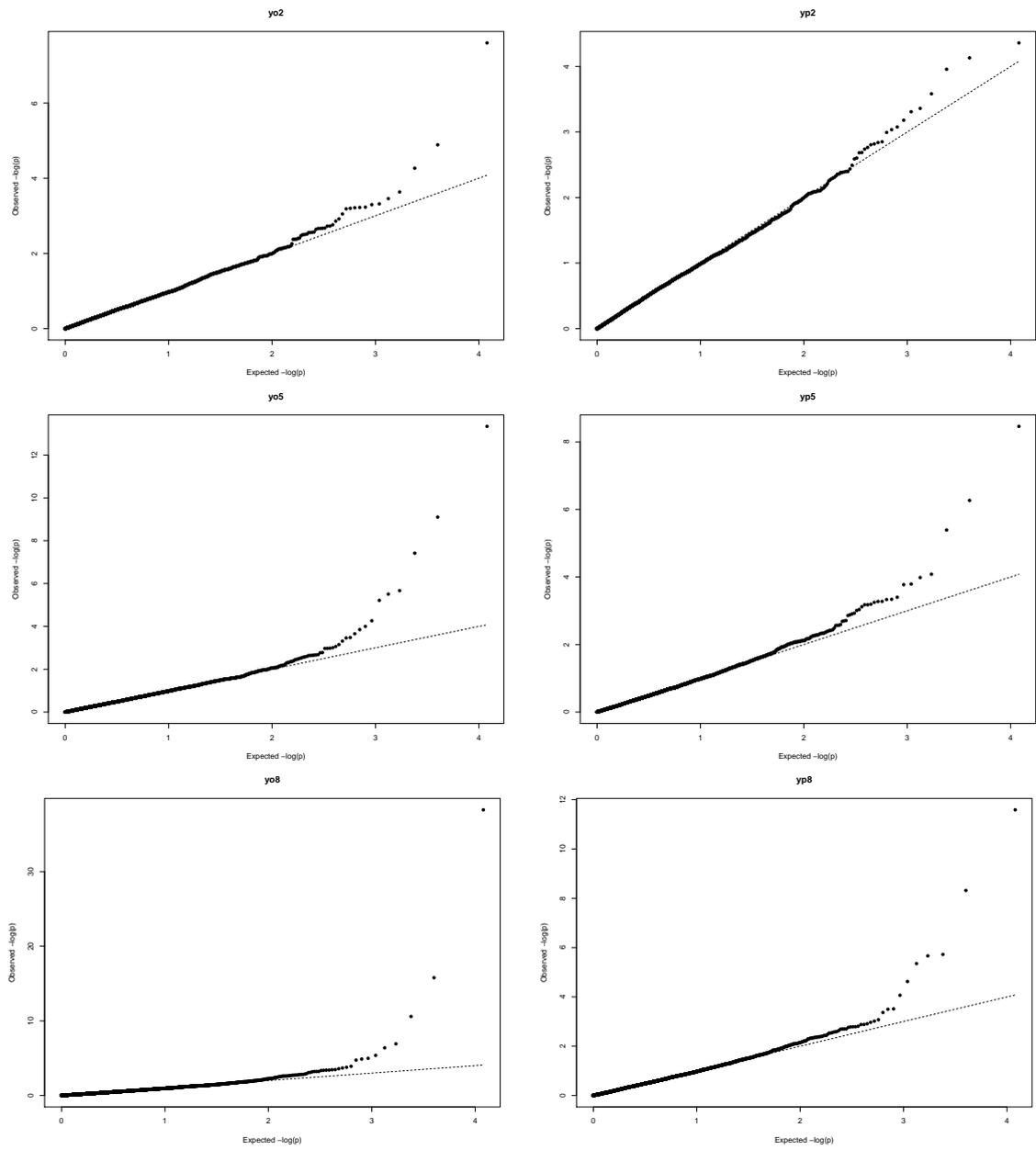
ANEXO B - GRÁFICOS EFEITOS SIMULADOS

Figura 1 – Estimativas da herdabilidade (h^2), variância genética (σ_a^2) e variância residual (σ_e^2) para vetores observação simulados em GWAS nos cenários oligogênico e poligênico pelo pacote GAPIT



Fonte: Da autora (2019)

Figura 2 – Gráficos QQ-plot pelo pacote rrBLUP para GWAS nos cenários oligogênico e poligênico



Fonte: Da autora (2019)

ANEXO C - TABELAS EFEITOS SIMULADOS

Tabela 1 – Tabelas Efeitos Simulados

| Resumo das informações dos 60 efeitos simulados (genes de interesse) para cenário Poligênico | | | |
|--|------------|-----------|-----------------------|
| Marca | Cromossomo | Posição | Efeito médio marcador |
| 382 | 1 | 262715518 | 1,163980 |
| 555 | 1 | 353591485 | 0,892086 |
| 182 | 1 | 74579886 | 0,491992 |
| 745 | 1 | 480911633 | 0,137613 |
| 457 | 1 | 284407382 | 0,082863 |
| 791 | 1 | 511736511 | 0,127505 |
| 554 | 1 | 352921379 | 0,179842 |
| 389 | 1 | 264847349 | 0,242850 |
| 89 | 1 | 28084996 | 0,764190 |
| 1562 | 2 | 363489782 | 1,829971 |
| 1404 | 2 | 250585181 | 0,808478 |
| 1304 | 2 | 189293647 | 0,522426 |
| 1744 | 2 | 493544448 | 0,137744 |
| 1653 | 2 | 428517115 | 0,165315 |
| 1119 | 2 | 27403594 | 0,637734 |
| 1646 | 2 | 423515013 | 0,960402 |
| 1093 | 2 | 20954874 | 1,370626 |
| 2626 | 3 | 481639278 | 0,431786 |
| 2895 | 3 | 731873569 | 0,327602 |
| 2547 | 3 | 408150397 | 0,022850 |
| 2035 | 3 | 17184071 | 0,085722 |
| 2279 | 3 | 194643731 | 0,848928 |
| 2517 | 3 | 380243227 | 1,548215 |
| 3261 | 4 | 212893509 | 1,468956 |
| 3218 | 4 | 178735388 | 0,546690 |
| 3113 | 4 | 107751627 | 0,545053 |
| 3897 | 4 | 247534742 | 0,469648 |
| 3529 | 4 | 246005702 | 0,133544 |
| 3440 | 4 | 245635907 | 0,421981 |
| 3448 | 4 | 245669147 | 0,428376 |
| 3772 | 4 | 247015367 | 0,701117 |
| 3337 | 4 | 245207942 | 1,210165 |
| 3960 | 4 | 247796507 | 1,346749 |
| 3288 | 4 | 235779015 | 1,448958 |
| 3480 | 4 | 245802107 | 2,377477 |
| 4951 | 5 | 766550402 | 1,288210 |
| 4648 | 5 | 485934224 | 1,249249 |
| 4549 | 5 | 394247750 | 1,248718 |
| 4629 | 5 | 468337830 | 1,154861 |
| 4849 | 5 | 672085550 | 1,051124 |
| 4266 | 5 | 185596985 | 0,338568 |
| 4952 | 5 | 767476528 | 0,315934 |
| 4121 | 5 | 58576757 | 0,161316 |
| 4898 | 5 | 717465724 | 0,113686 |
| 4987 | 5 | 799890938 | 0,108768 |
| 4828 | 5 | 652636904 | 0,233315 |
| 4942 | 5 | 758215268 | 0,383644 |
| 4123 | 5 | 58616019 | 0,455439 |
| 5826 | 6 | 386200335 | 2,037308 |
| 5033 | 6 | 75187712 | 1,513478 |
| 5560 | 6 | 291465765 | 0,887959 |
| 5805 | 6 | 378721290 | 0,471411 |
| 5104 | 6 | 113384662 | 0,437593 |
| 5345 | 6 | 214894590 | 0,348086 |
| 5422 | 6 | 242317755 | 0,342757 |
| 5296 | 6 | 197443485 | 0,172632 |
| 5405 | 6 | 236263290 | 0,182614 |
| 5809 | 6 | 380145870 | 0,540255 |
| 5862 | 6 | 399021555 | 1,006820 |
| 5590 | 6 | 302150115 | 1,164618 |

| Resumo das informações dos seis efeitos simulados (genes de interesse) para cenário Oligogênico | | | |
|---|------------|-----------|-----------------------|
| Marca | Cromossomo | Posição | Efeito médio marcador |
| 538 | 1 | 298413091 | 0,398856 |
| 2000 | 2 | 676478485 | 0,248179 |
| 2088 | 3 | 69892547 | 0,315044 |
| 3677 | 4 | 246620642 | 0,856388 |
| 4004 | 5 | 246620642 | 1,335127 |
| 4877 | 5 | 698017078 | 1,179022 |

Fonte: Da autora (2019)

Fonte: Da autora (2019)

Tabelas RR-BLUP x RR-BLUP modificado

Tabela 2 – RR-BLUP × RR-BLUP modificado
cenário Oligogênico, $h^2 = 0,2$

| Informações dos SNPS com efeito médio dos marcadores obtidos pelo método RR-BLUP pacote rrBLUP do <i>software</i> R | | | | Informações dos SNPS com efeito médio dos marcadores obtidos pelo método RR-BLUP modificado | | | |
|---|------------|-----------|-----------------------|---|------------|-----------|-----------------------|
| Marca | Cromossomo | Posição | Efeito médio marcador | Marca | Cromossomo | Posição | Efeito médio marcador |
| 4004 | 5 | 2123914 | 0,022656341 | 4004 | 5 | 2123914 | 0,016448067 |
| 2983 | 3 | 813734601 | 0,015695962 | 1029 | 2 | 4679573 | 0,010517630 |
| 2897 | 3 | 733734047 | 0,014594873 | 2983 | 3 | 813734601 | 0,009790817 |
| 4303 | 5 | 200813183 | 0,014166221 | 5521 | 6 | 277576110 | 0,009489235 |
| 1029 | 2 | 4679573 | 0,013966825 | 2897 | 3 | 733734047 | 0,009423459 |
| 4621 | 5 | 460928822 | 0,013669151 | 4260 | 5 | 181672656 | 0,008940788 |
| 2896 | 3 | 732803808 | 0,013563216 | 4621 | 5 | 460928822 | 0,008803985 |
| 1841 | 2 | 562859298 | 0,013522921 | 353 | 1 | 240574247 | 0,008785314 |
| 4047 | 5 | 9965534 | 0,013260689 | 2909 | 3 | 744896915 | 0,008776953 |
| 5521 | 6 | 277576110 | 0,013222457 | 5028 | 6 | 57774238 | 0,008520623 |

marca grifada - associação com gene simulado

Fonte: Da autora (2019)

Tabela 3 – RR-BLUP × RR-BLUP modificado
cenário Oligogênico, $h^2 = 0,5$

| Informações dos SNPS com efeito médio dos marcadores obtidos pelo método RR-BLUP pacote rrBLUP do <i>software</i> R | | | | Informações dos SNPS com efeito médio dos marcadores obtidos pelo método RR-BLUP modificado | | | |
|---|------------|-----------|-----------------------|---|------------|-----------|-----------------------|
| Marca | Cromossomo | Posição | Efeito médio marcador | Marca | Cromossomo | Posição | Efeito médio marcador |
| 4004 | 5 | 2123914 | 0,03960630 | 4004 | 5 | 2123914 | 0,018899077 |
| 3677 | 4 | 246620642 | 0,03170586 | 3677 | 4 | 246620642 | 0,016063392 |
| 4877 | 5 | 698017078 | 0,02609724 | 4877 | 5 | 698017078 | 0,014628395 |
| 3805 | 4 | 247152482 | 0,02256555 | 3805 | 4 | 247152482 | 0,012565509 |
| 4625 | 5 | 464633326 | 0,02182781 | 5265 | 6 | 186402990 | 0,012334083 |
| 3933 | 4 | 247684322 | 0,02103676 | 3933 | 4 | 247684322 | 0,011826708 |
| 5265 | 6 | 186402990 | 0,01973824 | 4260 | 5 | 181672656 | 0,010819111 |
| 4260 | 5 | 181672656 | 0,01970392 | 3792 | 4 | 247098467 | 0,010788463 |
| 5009 | 6 | 14057156 | 0,01900065 | 4388 | 5 | 245141464 | 0,010676607 |
| 4772 | 5 | 600773848 | 0,01882034 | 4625 | 5 | 464633326 | 0,010594175 |

marca grifada - associação com gene simulado

Fonte: Da autora (2019)

Tabela 4 – RR-BLUP × RR-BLUP modificado
cenário Oligogênico, $h^2 = 0,8$

| Informações dos SNPS com efeito médio dos marcadores obtidos pelo método RR-BLUP pacote rrBLUP do <i>software</i> R | | | | Informações dos SNPS com efeito médio dos marcadores obtidos pelo método RR-BLUP modificado | | | |
|---|------------|-----------|-----------------------|---|------------|-----------|-----------------------|
| Marca | Cromossomo | Posição | Efeito médio marcador | Marca | Cromossomo | Posição | Efeito médio marcador |
| 4004 | 5 | 2123914 | 0,08446578 | 4004 | 5 | 2123914 | 0,030377543 |
| 4877 | 5 | 698017078 | 0,04854806 | 4877 | 5 | 698017078 | 0,017173805 |
| 4260 | 5 | 181672656 | 0,03632223 | 4303 | 5 | 200813183 | 0,015148100 |
| 3677 | 4 | 246620642 | 0,03460128 | 4260 | 5 | 181672656 | 0,015074554 |
| 4303 | 5 | 200813183 | 0,03273183 | 3677 | 4 | 246620642 | 0,014564602 |
| 4237 | 5 | 171688991 | 0,02789022 | 4772 | 5 | 600773848 | 0,012838414 |
| 4047 | 5 | 9965534 | 0,02785691 | 5265 | 6 | 186402990 | 0,012662388 |
| 5412 | 6 | 238756305 | 0,02713492 | 4388 | 5 | 245141464 | 0,012308095 |
| 4625 | 5 | 464633326 | 0,02692567 | 4621 | 5 | 460928822 | 0,012125321 |
| 538 | 1 | 298413091 | 0,02448266 | 4047 | 5 | 9965534 | 0,012104269 |

marca grifada - associação com gene simulado

Fonte: Da autora (2019)

Tabela 5 – RR-BLUP × RR-BLUP modificado
cenário Poligênico, $h^2 = 0,2$

| Informações dos SNPS com efeito médio dos marcadores obtidos pelo RR-BLUP pacote rrBLUP <i>software</i> R | | | | Informações dos SNPS com efeito médio dos marcadores obtidos pelo RR-BLUP modificado | | | |
|---|------------|-----------|-----------------------|--|------------|-----------|-----------------------|
| Marca | Cromossomo | Posição | Efeito médio marcador | Marca | Cromossomo | Posição | Efeito médio marcador |
| 5826 | 6 | 386200335 | 0,05105054 | 5862 | 6 | 399021555 | 0,03255141 |
| 5862 | 6 | 399021555 | 0,04988037 | 5826 | 6 | 386200335 | 0,02952465 |
| 3941 | 4 | 247717562 | 0,04894269 | 5990 | 6 | 444608115 | 0,02772593 |
| 3480 | 4 | 245802107 | 0,04704041 | 3480 | 4 | 245802107 | 0,02644800 |
| 4966 | 5 | 780442292 | 0,04696129 | 3941 | 4 | 247717562 | 0,02633346 |
| 5801 | 6 | 377296710 | 0,04339915 | 5269 | 6 | 187827570 | 0,02567233 |
| 3658 | 4 | 246541697 | 0,04158533 | 3766 | 4 | 246990437 | 0,02537274 |
| 3517 | 4 | 245955842 | 0,04135003 | 5239 | 6 | 177143220 | 0,02492467 |
| 805 | 1 | 521117995 | 0,04095748 | 4966 | 5 | 780442292 | 0,02433416 |
| 1408 | 2 | 253443525 | 0,04093025 | 3658 | 4 | 246541697 | 0,02415166 |
| 3766 | 4 | 246990437 | 0,03972687 | 5801 | 6 | 377296710 | 0,02387703 |
| 1573 | 2 | 371350229 | 0,03967739 | 3718 | 4 | 246790997 | 0,02361566 |
| 3718 | 4 | 246790997 | 0,03922396 | 4070 | 5 | 13981667 | 0,02295932 |
| 5239 | 6 | 177143220 | 0,03888178 | 5728 | 6 | 351298125 | 0,02294928 |
| 2840 | 3 | 680710424 | 0,03883639 | 3126 | 4 | 122795391 | 0,02265420 |
| 4070 | 5 | 13981667 | 0,03818778 | 3517 | 4 | 245955842 | 0,02230773 |
| 5763 | 6 | 363763200 | 0,03809618 | 1573 | 2 | 371350229 | 0,02228555 |
| 1793 | 2 | 528559166 | 0,03773059 | 4960 | 5 | 774885536 | 0,02227711 |
| 5990 | 6 | 444608115 | 0,03767349 | 5478 | 6 | 262261875 | 0,02219500 |
| 5728 | 6 | 351298125 | 0,03749504 | 5763 | 6 | 363763200 | 0,02191691 |
| 3515 | 4 | 245947532 | 0,03671659 | 4582 | 5 | 424809908 | 0,02177673 |
| 1941 | 2 | 634317906 | 0,03667907 | 3044 | 4 | 35384247 | 0,02133740 |
| 5478 | 6 | 262261875 | 0,03661736 | 5350 | 6 | 216675315 | 0,02130933 |
| 5269 | 6 | 187827570 | 0,03619892 | 1408 | 2 | 253443525 | 0,02122744 |
| 4838 | 5 | 661898164 | 0,03584389 | 842 | 1 | 545911919 | 0,02108450 |
| 1268 | 2 | 169989408 | 0,03580397 | 37 | 1 | 9300541 | 0,02105339 |
| 3960 | 4 | 247796507 | 0,03567401 | 989 | 1 | 644417507 | 0,02099247 |
| 3810 | 4 | 247173257 | 0,03520935 | 2616 | 3 | 472336888 | 0,02082974 |
| 4582 | 5 | 424809908 | 0,03516500 | 5673 | 6 | 331710150 | 0,02065441 |
| 5350 | 6 | 216675315 | 0,03496114 | 4838 | 5 | 661898164 | 0,02064937 |
| 5123 | 6 | 121839487 | 0,03436159 | 699 | 1 | 450086755 | 0,02055844 |
| 5789 | 6 | 373022970 | 0,03434489 | 4862 | 5 | 684125188 | 0,02050635 |
| 842 | 1 | 545911919 | 0,03428186 | 3964 | 4 | 247813127 | 0,02016463 |
| 3420 | 4 | 245552807 | 0,03388879 | 629 | 1 | 403179332 | 0,02011990 |
| 3044 | 4 | 35384247 | 0,03365675 | 1268 | 2 | 169989408 | 0,01996070 |
| 3126 | 4 | 122795391 | 0,03349254 | 933 | 1 | 606891568 | 0,01989972 |
| 2101 | 3 | 110715956 | 0,03328573 | 1935 | 2 | 630030390 | 0,01986214 |
| 3780 | 4 | 247048607 | 0,03321652 | 3354 | 4 | 245278577 | 0,01983601 |
| 1617 | 2 | 402792016 | 0,03312781 | 1600 | 2 | 390644053 | 0,01978147 |
| 382 | 1 | 262715518 | 0,03299309 | 2541 | 3 | 402568963 | 0,01967526 |
| 3881 | 4 | 247468262 | 0,03294294 | 3881 | 4 | 247468262 | 0,01967216 |
| 1865 | 2 | 580009364 | 0,03287584 | 5818 | 6 | 383351175 | 0,01963770 |
| 5379 | 6 | 227003520 | 0,03276607 | 5789 | 6 | 373022970 | 0,01963465 |
| 2578 | 3 | 436987806 | 0,03251639 | 1617 | 2 | 402792016 | 0,01948798 |
| 1935 | 2 | 630030390 | 0,03231497 | 2578 | 3 | 436987806 | 0,01947892 |
| 629 | 1 | 403179332 | 0,03218485 | 4403 | 5 | 259033354 | 0,01947437 |
| 4862 | 5 | 684125188 | 0,03208060 | 5379 | 6 | 227003520 | 0,01926845 |
| 2726 | 3 | 574663178 | 0,03202130 | 3164 | 4 | 149588842 | 0,01923991 |
| 3296 | 4 | 239237605 | 0,03190238 | 4961 | 5 | 775811662 | 0,01912766 |
| 5277 | 6 | 190676730 | 0,03187561 | 3396 | 4 | 245453087 | 0,01912126 |
| 5693 | 6 | 338833050 | 0,03179861 | 3885 | 4 | 247484882 | 0,01907119 |
| 4365 | 5 | 223840566 | 0,03174967 | 3454 | 4 | 245694077 | 0,01905919 |
| 4960 | 5 | 774885536 | 0,03174620 | 1562 | 2 | 363489782 | 0,01902395 |
| 5954 | 6 | 431786895 | 0,03167186 | 825 | 1 | 534520116 | 0,01901383 |
| 3354 | 4 | 245278577 | 0,03163290 | 5277 | 6 | 190676730 | 0,01899574 |
| 5933 | 6 | 424307850 | 0,03162938 | 3515 | 4 | 245947532 | 0,01883465 |
| 1657 | 2 | 431375459 | 0,03148339 | 3377 | 4 | 245374142 | 0,01880962 |
| 94 | 1 | 32583282 | 0,03146275 | 65 | 1 | 17601517 | 0,01875373 |
| 933 | 1 | 606891568 | 0,03133262 | 4023 | 5 | 6820571 | 0,01868914 |
| 794 | 1 | 513746829 | 0,03124857 | 2369 | 3 | 242567855 | 0,01864365 |
| 3940 | 4 | 247713407 | 0,03099296 | 5830 | 6 | 387624915 | 0,01861624 |

marca grifada - associação com gene simulado

Fonte: Da autora (2019)

Tabela 6 – RR-BLUP × RR-BLUP modificado
cenário Poligênico, $h^2 = 0,5$

| Informações dos SNPs com efeito médio dos marcadores obtidos pelo método RR-BLUP pacote rrBLUP <i>software</i> R | | | | Informações dos SNPs com efeito médio dos marcadores obtidos pelo método RR-BLUP modificado | | | |
|--|------------|-----------|-----------------------|---|------------|-----------|-----------------------|
| Marca | Cromossomo | Posição | Efeito médio marcador | Marca | Cromossomo | Posição | Efeito médio marcador |
| 5826 | 6 | 386200335 | 0,09946467 | 5826 | 6 | 386200335 | 0,04575881 |
| 3480 | 4 | 245802107 | 0,07872885 | 3480 | 4 | 245802107 | 0,03774874 |
| 5846 | 6 | 393323235 | 0,06988989 | 5673 | 6 | 331710150 | 0,03700575 |
| 1646 | 2 | 423515013 | 0,06561643 | 3517 | 4 | 245955842 | 0,03528487 |
| 5673 | 6 | 331710150 | 0,06448909 | 5277 | 6 | 190676730 | 0,03422082 |
| 1562 | 2 | 363489782 | 0,06402292 | 5862 | 6 | 399021555 | 0,03391611 |
| 3261 | 4 | 212893509 | 0,06321896 | 5846 | 6 | 393323235 | 0,03217211 |
| 3517 | 4 | 245955842 | 0,06305742 | 5884 | 6 | 406856745 | 0,03089208 |
| 5862 | 6 | 399021555 | 0,06030942 | 4337 | 5 | 209498142 | 0,03067210 |
| 4317 | 5 | 204516414 | 0,05942433 | 3261 | 4 | 212893509 | 0,03032911 |
| 1093 | 2 | 20954874 | 0,05883710 | 1562 | 2 | 363489782 | 0,03028617 |
| 5884 | 6 | 406856745 | 0,05816223 | 4838 | 5 | 661898164 | 0,02914228 |
| 5277 | 6 | 190676730 | 0,05792280 | 4317 | 5 | 204516414 | 0,02890251 |
| 1865 | 2 | 580009364 | 0,05637206 | 5763 | 6 | 363763200 | 0,02823753 |
| 4649 | 5 | 486860350 | 0,05472832 | 1093 | 2 | 20954874 | 0,02737217 |
| 1185 | 2 | 58354695 | 0,05331042 | 4290 | 5 | 193521948 | 0,02723381 |
| 5933 | 6 | 424307850 | 0,05329053 | 4649 | 5 | 486860350 | 0,02720298 |
| 187 | 1 | 83595286 | 0,05323705 | 5619 | 6 | 312478320 | 0,02690338 |
| 5805 | 6 | 378721290 | 0,05251513 | 1865 | 2 | 580009364 | 0,02661258 |
| 2517 | 3 | 380243227 | 0,05213519 | 1646 | 2 | 423515013 | 0,02639495 |
| 4838 | 5 | 661898164 | 0,05183724 | 4977 | 5 | 790629678 | 0,02621186 |
| 945 | 1 | 614932841 | 0,05144960 | 4913 | 5 | 731357614 | 0,02553390 |
| 2749 | 3 | 596058675 | 0,05083807 | 2749 | 3 | 596058675 | 0,02552698 |
| 1354 | 2 | 214263400 | 0,05040887 | 2072 | 3 | 53549251 | 0,02531463 |
| 4265 | 5 | 184588958 | 0,04992513 | 4265 | 5 | 184588958 | 0,02530047 |
| 4290 | 5 | 193521948 | 0,04920688 | 4966 | 5 | 780442292 | 0,02523692 |
| 3389 | 4 | 245424002 | 0,04864611 | 3133 | 4 | 128632208 | 0,02510840 |
| 5619 | 6 | 312478320 | 0,04862274 | 5442 | 6 | 249440655 | 0,02504247 |
| 308 | 1 | 216989434 | 0,04853211 | 5547 | 6 | 286835880 | 0,02503977 |
| 4977 | 5 | 790629678 | 0,04833549 | 2602 | 3 | 459313542 | 0,02494441 |
| 5789 | 6 | 373022970 | 0,04809904 | 5244 | 6 | 178923945 | 0,02493813 |
| 5799 | 6 | 376584420 | 0,04794439 | 5853 | 6 | 395816250 | 0,02488867 |
| 5033 | 6 | 75187712 | 0,04790453 | 3992 | 4 | 247929467 | 0,02484166 |
| 4170 | 5 | 75234294 | 0,04785791 | 5698 | 6 | 340613775 | 0,02479061 |
| 1358 | 2 | 214423504 | 0,04763065 | 4453 | 5 | 305339654 | 0,02446541 |
| 5698 | 6 | 340613775 | 0,04754138 | 4126 | 5 | 59300236 | 0,02439315 |
| 499 | 1 | 291031959 | 0,04742413 | 945 | 1 | 614932841 | 0,02426242 |
| 1955 | 2 | 644322111 | 0,04737888 | 2517 | 3 | 380243227 | 0,02425399 |
| 1321 | 2 | 198757428 | 0,04713065 | 3009 | 4 | 2639015 | 0,02422617 |
| 5763 | 6 | 363763200 | 0,04700885 | 4081 | 5 | 199733330 | 0,02397715 |
| 4337 | 5 | 209498142 | 0,04681282 | 5033 | 6 | 75187712 | 0,02370273 |
| 1338 | 2 | 206974778 | 0,04678307 | 5361 | 6 | 220592910 | 0,02354980 |
| 1210 | 2 | 128180134 | 0,04675315 | 3633 | 4 | 246437822 | 0,02350999 |
| 4453 | 5 | 305339654 | 0,04671432 | 499 | 1 | 291031959 | 0,02349486 |
| 427 | 1 | 268373293 | 0,04664152 | 4170 | 5 | 75234294 | 0,02348769 |
| 3972 | 4 | 247846367 | 0,04639436 | 3224 | 4 | 184373943 | 0,02334718 |
| 89 | 1 | 28084996 | 0,04612228 | 5799 | 6 | 376584420 | 0,02332370 |
| 2877 | 3 | 715129267 | 0,04611050 | 1210 | 2 | 128180134 | 0,02318648 |
| 3992 | 4 | 247929467 | 0,04590856 | 1417 | 2 | 259874800 | 0,02317542 |
| 1417 | 2 | 259874800 | 0,04588461 | 3864 | 4 | 247397627 | 0,02308143 |
| 4913 | 5 | 731357614 | 0,04559348 | 5933 | 6 | 424307850 | 0,02277039 |
| 1409 | 2 | 254158112 | 0,04531763 | 427 | 1 | 268373293 | 0,02270288 |
| 3009 | 4 | 2639015 | 0,04527754 | 4554 | 5 | 398878380 | 0,02264047 |
| 4629 | 5 | 468337830 | 0,04526160 | 1354 | 2 | 214263400 | 0,02260491 |
| 1550 | 2 | 354914749 | 0,04502627 | 1263 | 2 | 169265278 | 0,02247365 |
| 821 | 1 | 531839692 | 0,04492840 | 1358 | 2 | 214423504 | 0,02245957 |
| 3133 | 4 | 128632208 | 0,04491735 | 1185 | 2 | 58354695 | 0,02236117 |
| 1313 | 2 | 195555350 | 0,04480449 | 3960 | 4 | 247796507 | 0,02230172 |
| 2045 | 3 | 27199028 | 0,04458147 | 1631 | 2 | 412796221 | 0,02223979 |
| 5461 | 6 | 256207410 | 0,04446594 | 2746 | 3 | 593267958 | 0,02223056 |
| 2072 | 3 | 53549251 | 0,04443730 | 3389 | 4 | 245424002 | 0,02222397 |
| 514 | 1 | 293632633 | 0,04439467 | 5085 | 6 | 91883155 | 0,02221634 |

marca grafada - associação com gene simulado

Fonte: Da autora (2019)

Tabela 7 – RR-BLUP × RR-BLUP modificado
cenário Poligênico, $h^2 = 0,8$

| Informações dos SNPS com efeito médio dos marcadores obtidos pelo método RR-BLUP pacote rrBLUP <i>software</i> R | | | | Informações dos SNPS com efeito médio dos marcadores obtidos pelo método RR-BLUP modificado | | | |
|--|------------|-----------|-----------------------|---|------------|-----------|-----------------------|
| Marca | Cromossomo | Posição | Efeito médio marcador | Marca | Cromossomo | Posição | Efeito médio marcador |
| 5826 | 6 | 386200335 | 0,16619728 | 5826 | 6 | 386200335 | 0,05659110 |
| 3480 | 4 | 245802107 | 0,12717157 | 3480 | 4 | 245802107 | 0,03851578 |
| 5954 | 6 | 431786895 | 0,10746622 | 1562 | 2 | 363489782 | 0,03707781 |
| 5862 | 6 | 399021555 | 0,10413423 | 3517 | 4 | 245955842 | 0,03500640 |
| 5033 | 6 | 75187712 | 0,09963325 | 5862 | 6 | 399021555 | 0,03500279 |
| 1562 | 2 | 363489782 | 0,09403024 | 5954 | 6 | 431786895 | 0,03371941 |
| 1093 | 2 | 20954874 | 0,09031506 | 5033 | 6 | 75187712 | 0,03060106 |
| 5590 | 6 | 302150115 | 0,08362926 | 3658 | 4 | 246541697 | 0,03035686 |
| 89 | 1 | 28084996 | 0,08297049 | 5763 | 6 | 363763200 | 0,03020212 |
| 2517 | 3 | 380243227 | 0,08080153 | 5801 | 6 | 377296710 | 0,02964773 |
| 1646 | 2 | 423515013 | 0,08041383 | 3998 | 4 | 247954397 | 0,02900643 |
| 859 | 1 | 557303721 | 0,07898176 | 4649 | 5 | 486860350 | 0,02894331 |
| 1865 | 2 | 580009364 | 0,07846753 | 5846 | 6 | 393323235 | 0,02885648 |
| 3517 | 4 | 245955842 | 0,07681823 | 3261 | 4 | 212893509 | 0,02860739 |
| 3918 | 4 | 247621997 | 0,07603372 | 1646 | 2 | 423515013 | 0,02849573 |
| 966 | 1 | 629005068 | 0,07588504 | 1865 | 2 | 580009364 | 0,02841352 |
| 2292 | 3 | 200201258 | 0,07562277 | 4629 | 5 | 468337830 | 0,02836645 |
| 3996 | 4 | 247946087 | 0,07551852 | 3810 | 4 | 247173257 | 0,02810733 |
| 5805 | 6 | 378721290 | 0,07467317 | 2517 | 3 | 380243227 | 0,02808315 |
| 5134 | 6 | 131576099 | 0,07446444 | 1686 | 2 | 452098456 | 0,02765476 |
| 3643 | 4 | 246479372 | 0,07443420 | 3996 | 4 | 247946087 | 0,02755433 |
| 1086 | 2 | 19837545 | 0,07293890 | 966 | 1 | 629005068 | 0,02744267 |
| 1686 | 2 | 452098456 | 0,07276278 | 5442 | 6 | 249440655 | 0,02735728 |
| 5782 | 6 | 370529955 | 0,07245163 | 3918 | 4 | 247621997 | 0,02685934 |
| 1880 | 2 | 590728155 | 0,07212221 | 3760 | 4 | 246965507 | 0,02678381 |
| 5636 | 6 | 318532785 | 0,07191140 | 3643 | 4 | 246479372 | 0,02676693 |
| 4629 | 5 | 468337830 | 0,07150437 | 382 | 1 | 262715518 | 0,02668548 |
| 5801 | 6 | 377296710 | 0,07084332 | 5698 | 6 | 340613775 | 0,02666688 |
| 1249 | 2 | 161879250 | 0,07071994 | 5774 | 6 | 367680795 | 0,02655562 |
| 3658 | 4 | 246541697 | 0,07060408 | 5673 | 6 | 331710150 | 0,02655351 |
| 3476 | 4 | 245785487 | 0,06923100 | 3220 | 4 | 180309373 | 0,02646546 |
| 5530 | 6 | 280781415 | 0,06886680 | 5171 | 6 | 149251173 | 0,02622642 |
| 1304 | 2 | 189293647 | 0,06879842 | 2901 | 3 | 737455003 | 0,02621439 |
| 728 | 1 | 469519830 | 0,06832792 | 5186 | 6 | 157129197 | 0,02613092 |
| 382 | 1 | 262715518 | 0,06789890 | 4265 | 5 | 184588958 | 0,02596787 |
| 1178 | 2 | 51785620 | 0,06785275 | 3389 | 4 | 245424002 | 0,02567719 |
| 5846 | 6 | 393323235 | 0,06735277 | 4838 | 5 | 661898164 | 0,02546067 |
| 5982 | 6 | 441758955 | 0,06717276 | 1178 | 2 | 51785620 | 0,02538399 |
| 2273 | 3 | 193565736 | 0,06699139 | 5882 | 6 | 406144455 | 0,02535180 |
| 2998 | 3 | 827688186 | 0,06665137 | 5357 | 6 | 219168330 | 0,02530312 |
| 5290 | 6 | 195306615 | 0,06660403 | 5884 | 6 | 406856745 | 0,02513824 |
| 5882 | 6 | 406144455 | 0,06641910 | 2840 | 3 | 680710424 | 0,02501926 |
| 1816 | 2 | 544994646 | 0,06628584 | 5853 | 6 | 395816250 | 0,02495355 |
| 4849 | 5 | 672085550 | 0,06617514 | 5683 | 6 | 335271600 | 0,02472734 |
| 1917 | 2 | 617167840 | 0,06594504 | 2998 | 3 | 827688186 | 0,02471365 |
| 6000 | 6 | 448169565 | 0,06514818 | 5619 | 6 | 312478320 | 0,02467295 |
| 2840 | 3 | 680710424 | 0,06493658 | 3301 | 4 | 242019500 | 0,02453895 |
| 5698 | 6 | 340613775 | 0,06476278 | 3515 | 4 | 245947532 | 0,02450517 |
| 3220 | 4 | 180309373 | 0,06448251 | 859 | 1 | 557303721 | 0,02450508 |
| 5186 | 6 | 157129197 | 0,06442244 | 1079 | 2 | 14751136 | 0,02448790 |
| 486 | 1 | 286642725 | 0,06421748 | 1661 | 2 | 434233804 | 0,02446436 |
| 4459 | 5 | 310896410 | 0,06412894 | 5547 | 6 | 286835880 | 0,02445658 |
| 3736 | 4 | 246865787 | 0,06410627 | 5982 | 6 | 441758955 | 0,02445053 |
| 3538 | 4 | 246043097 | 0,06401348 | 4476 | 5 | 326640552 | 0,02437252 |
| 3998 | 4 | 247954397 | 0,06387625 | 1759 | 2 | 504263240 | 0,02423941 |
| 1910 | 2 | 612165738 | 0,06382339 | 3476 | 4 | 245785487 | 0,02418216 |
| 5811 | 6 | 380858160 | 0,06378827 | 89 | 1 | 28084996 | 0,02411012 |
| 158 | 1 | 56175806 | 0,06361935 | 5811 | 6 | 380858160 | 0,02380398 |
| 1867 | 2 | 581438536 | 0,06347538 | 4735 | 5 | 566507186 | 0,02347495 |
| 4394 | 5 | 250698220 | 0,06344617 | 5134 | 6 | 131576099 | 0,02347286 |
| 5416 | 6 | 240180885 | 0,06340593 | 821 | 1 | 531839692 | 0,02343574 |
| 5171 | 6 | 149251173 | 0,06318698 | 3835 | 4 | 247277132 | 0,02342072 |

marca grifada - associação com gene simulado

Fonte: Da autora (2019)

Tabelas Bayes A x Bayes A modificado

Tabela 8 – Bayes A × Bayes A modificado
cenário Oligogênico, $h^2 = 0,2$

| Informações dos SNPS com efeito médio dos marcadores obtidos pelo método Bayes A pacote BGLR | | | | Informações dos SNPS com efeito médio dos marcadores obtidos pelo método Bayes A modificado | | | |
|--|------------|-----------|-----------------------|---|------------|-----------|-----------------------|
| Marca | Cromossomo | Posição | Efeito médio marcador | Marca | Cromossomo | Posição | Efeito médio marcador |
| 4004 | 5 | 2123914 | 0,05264070 | 4004 | 5 | 2123914 | 1,43928751 |
| 2983 | 3 | 813734601 | 0,03768931 | 2983 | 3 | 813734601 | 0,33470226 |
| 4047 | 5 | 9965534 | 0,03032409 | 1029 | 2 | 4679573 | 0,12705359 |
| 279 | 1 | 202620610 | 0,02934032 | 1826 | 2 | 552140507 | 0,09154366 |
| 922 | 1 | 599520402 | 0,02929689 | 4877 | 5 | 698017078 | 0,08643809 |
| 4980 | 5 | 793408056 | 0,02857487 | 279 | 1 | 202620610 | 0,08411740 |
| 1841 | 2 | 562859298 | 0,02849086 | 2256 | 3 | 186580149 | 0,07463366 |
| 4303 | 5 | 200813183 | 0,02847216 | 2897 | 3 | 733734047 | 0,07274871 |
| 4491 | 5 | 340532442 | 0,02585890 | 1051 | 2 | 7482709 | 0,06688483 |
| 3239 | 4 | 197964404 | 0,02459472 | 2896 | 3 | 732803808 | 0,04432225 |

marca grifada - associação com gene simulado

Fonte: Da autora (2019)

Tabela 9 – Bayes A × Bayes A modificado
cenário Oligogênico, $h^2 = 0,5$

| Informações dos SNPS com efeito médio dos marcadores obtidos pelo método Bayes A pacote BGLR | | | | Informações dos SNPS com efeito médio dos marcadores obtidos pelo método Bayes A modificado | | | |
|--|------------|-----------|-----------------------|---|------------|-----------|-----------------------|
| Marca | Cromossomo | Posição | Efeito médio marcador | Marca | Cromossomo | Posição | Efeito médio marcador |
| 4004 | 5 | 2123914 | 1,09213635 | 4004 | 5 | 2123914 | 1,178316079 |
| 3677 | 4 | 246620642 | 0,84730861 | 4877 | 5 | 698017078 | 1,064273408 |
| 4877 | 5 | 698017078 | 0,41460902 | 3677 | 4 | 246620642 | 1,059949819 |
| 625 | 1 | 400498908 | 0,02374197 | 193 | 1 | 88380847 | 0,103085058 |
| 1629 | 2 | 411367049 | 0,02340164 | 833 | 1 | 539880964 | 0,032716537 |
| 466 | 1 | 285364244 | 0,02263857 | 877 | 1 | 569365630 | 0,029506181 |
| 610 | 1 | 390447317 | 0,02219050 | 186 | 1 | 82563962 | 0,026207086 |
| 2081 | 3 | 64730607 | 0,02210578 | 538 | 1 | 298413091 | 0,025817553 |
| 117 | 1 | 41505211 | 0,02170303 | 625 | 1 | 400498908 | 0,025292679 |
| 193 | 1 | 88380847 | 0,02102575 | 713 | 1 | 459468239 | 0,021346092 |

marca grifada - associação com gene simulado

Fonte: Da autora (2019)

Tabela 10 – Bayes A × Bayes A modificado
cenário Oligogênico, $h^2 = 0,8$

| Informações dos SNPS com efeito médio dos marcadores obtidos pelo método Bayes A pacote BGLR | | | | Informações dos SNPS com efeito médio dos marcadores obtidos pelo método Bayes A modificado | | | |
|--|------------|-----------|-----------------------|---|------------|-----------|-----------------------|
| Marca | Cromossomo | Posição | Efeito médio marcador | Marca | Cromossomo | Posição | Efeito médio marcador |
| 4004 | 5 | 2123914 | 1,411451383 | 4004 | 5 | 2123914 | 1,424683336 |
| 4877 | 5 | 698017078 | 1,057546690 | 4877 | 5 | 698017078 | 1,081501332 |
| 3677 | 4 | 246620642 | 0,612117118 | 3677 | 4 | 246620642 | 0,718905761 |
| 538 | 1 | 298413091 | 0,042185384 | 538 | 1 | 298413091 | 0,218284727 |
| 2088 | 3 | 69892547 | 0,025319673 | 2088 | 3 | 69892547 | 0,132238418 |
| 5959 | 6 | 433567620 | 0,021230498 | 2000 | 2 | 676478485 | 0,036493772 |
| 189 | 1 | 87910951 | 0,015060302 | 4927 | 5 | 744323378 | 0,023602546 |
| 2094 | 3 | 108233038 | 0,014579287 | 2803 | 3 | 646291581 | 0,019298185 |
| 4928 | 5 | 745249504 | 0,013350015 | 3315 | 4 | 245051599 | 0,018056555 |
| 900 | 1 | 584778069 | 0,013251556 | 2559 | 3 | 419313265 | 0,013904708 |

marca grifada - associação com gene simulado

Fonte: Da autora (2019)

Tabela 11 – Bayes A × Bayes A modificado
cenário Poligênico, $h^2 = 0,2$

| Informações dos SNPS com efeito médio dos marcadores obtidos pelo método Bayes A pacote BGLR | | | | Informações dos SNPS com efeito médio dos marcadores obtidos pelo método Bayes A modificado | | | |
|--|------------|-----------|-----------------------|---|------------|-----------|-----------------------|
| Marca | Cromossomo | Posição | Efeito médio marcador | Marca | Cromossomo | Posição | Efeito médio marcador |
| 5862 | 6 | 399021555 | 0,11710371 | 5826 | 6 | 386200335 | 1,24447735 |
| 5826 | 6 | 386200335 | 0,11237839 | 3941 | 4 | 247717562 | 0,68905644 |
| 3941 | 4 | 247717562 | 0,11201358 | 3480 | 4 | 245802107 | 0,63353822 |
| 1408 | 2 | 253443525 | 0,09061189 | 5862 | 6 | 399021555 | 0,32367835 |
| 3480 | 4 | 245802107 | 0,08296427 | 2840 | 3 | 680710424 | 0,25611247 |
| 1941 | 2 | 634317906 | 0,08206468 | 5269 | 6 | 187827570 | 0,25510713 |
| 5801 | 6 | 377296710 | 0,08202645 | 1408 | 2 | 253443525 | 0,24316792 |
| 3462 | 4 | 245727317 | 0,07745465 | 3718 | 4 | 246790997 | 0,20964121 |
| 4966 | 5 | 780442292 | 0,07705477 | 1941 | 2 | 634317906 | 0,15666541 |
| 3718 | 4 | 246790997 | 0,07606185 | 2025 | 3 | 9124190 | 0,15518341 |
| 5239 | 6 | 177143220 | 0,07566884 | 3420 | 4 | 245552807 | 0,15459502 |
| 3766 | 4 | 246990437 | 0,07457886 | 5801 | 6 | 377296710 | 0,15296668 |
| 2840 | 3 | 680710424 | 0,07333674 | 3515 | 4 | 245947532 | 0,14435613 |
| 4606 | 5 | 447036932 | 0,07270704 | 1562 | 2 | 363489782 | 0,13502734 |
| 4862 | 5 | 684125188 | 0,07233228 | 5239 | 6 | 177143220 | 0,13221797 |
| 1793 | 2 | 528559166 | 0,07209929 | 5123 | 6 | 121839487 | 0,12644663 |
| 745 | 1 | 480911633 | 0,07189087 | 805 | 1 | 521117995 | 0,12541121 |
| 65 | 1 | 17601517 | 0,06912330 | 1573 | 2 | 371350229 | 0,12499424 |
| 5789 | 6 | 373022970 | 0,06834486 | 3810 | 4 | 247173257 | 0,11428329 |
| 5990 | 6 | 444608115 | 0,06769908 | 3296 | 4 | 239237605 | 0,11015044 |
| 1570 | 2 | 369206470 | 0,06747811 | 3960 | 4 | 247796507 | 0,10993435 |
| 2578 | 3 | 436987806 | 0,06719819 | 3766 | 4 | 246990437 | 0,10987716 |
| 5277 | 6 | 190676730 | 0,06654135 | 3517 | 4 | 245955842 | 0,10669571 |
| 37 | 1 | 9300541 | 0,06555347 | 3044 | 4 | 35384247 | 0,09450092 |
| 842 | 1 | 545911919 | 0,06499504 | 4365 | 5 | 223840566 | 0,09235704 |
| 3658 | 4 | 246541697 | 0,06424728 | 5728 | 6 | 351298125 | 0,08904939 |
| 2890 | 3 | 727222374 | 0,06365057 | 3399 | 4 | 245465552 | 0,08827677 |
| 3517 | 4 | 245955842 | 0,06329197 | 5478 | 6 | 262261875 | 0,08422467 |
| 805 | 1 | 521117995 | 0,06318111 | 4795 | 5 | 622074746 | 0,08289909 |
| 4838 | 5 | 661898164 | 0,06292388 | 5357 | 6 | 219168330 | 0,08285662 |
| 5763 | 6 | 363763200 | 0,06282043 | 1268 | 2 | 169989408 | 0,08077281 |
| 1076 | 2 | 13242487 | 0,06268772 | 5034 | 6 | 78304583 | 0,08070761 |
| 4087 | 5 | 21770965 | 0,06265669 | 382 | 1 | 262715518 | 0,07985038 |
| 5756 | 6 | 361270185 | 0,06257922 | 3354 | 4 | 245278577 | 0,07951357 |
| 1804 | 2 | 536419613 | 0,06252548 | 3963 | 4 | 247808972 | 0,07909555 |
| 3515 | 4 | 245947532 | 0,06240381 | 5990 | 6 | 444608115 | 0,07682056 |
| 3846 | 4 | 247322837 | 0,06219993 | 4070 | 5 | 13981667 | 0,07617753 |
| 146 | 1 | 50197644 | 0,06214839 | 4966 | 5 | 780442292 | 0,07581414 |
| 4365 | 5 | 223840566 | 0,06210607 | 2613 | 3 | 469546171 | 0,07451917 |
| 3865 | 4 | 247401782 | 0,06090028 | 745 | 1 | 480911633 | 0,07448195 |
| 5034 | 6 | 78304583 | 0,06072570 | 3126 | 4 | 122795391 | 0,07447921 |
| 5728 | 6 | 351298125 | 0,06072309 | 2480 | 3 | 345824384 | 0,07329528 |
| 4746 | 5 | 576694572 | 0,06046022 | 4682 | 5 | 517422508 | 0,07173399 |
| 1173 | 2 | 51780601 | 0,06044406 | 4087 | 5 | 21770965 | 0,06907266 |
| 933 | 1 | 606891568 | 0,05993441 | 5693 | 6 | 338833050 | 0,06888742 |
| 5317 | 6 | 204922530 | 0,05966216 | 5175 | 6 | 151958886 | 0,06869805 |
| 1402 | 2 | 249156009 | 0,05918449 | 4582 | 5 | 424809908 | 0,06817323 |
| 629 | 1 | 403179332 | 0,05914168 | 1657 | 2 | 431375459 | 0,06787237 |
| 1573 | 2 | 371350229 | 0,05904441 | 251 | 1 | 183647544 | 0,06755433 |
| 2117 | 3 | 122632201 | 0,05871413 | 1617 | 2 | 402792016 | 0,06664794 |
| 2101 | 3 | 110715956 | 0,05844317 | 794 | 1 | 513746829 | 0,06489567 |
| 2596 | 3 | 453732108 | 0,05832853 | 842 | 1 | 545911919 | 0,06486067 |
| 1225 | 2 | 140264872 | 0,05829669 | 1225 | 2 | 140264872 | 0,06442011 |
| 969 | 1 | 631015386 | 0,05827802 | 3259 | 4 | 210236675 | 0,06201453 |
| 3780 | 4 | 247048607 | 0,05813058 | 2101 | 3 | 110715956 | 0,06040325 |
| 1705 | 2 | 465675591 | 0,05810233 | 933 | 1 | 606891568 | 0,05898085 |
| 1617 | 2 | 402792016 | 0,05767169 | 2393 | 3 | 264893591 | 0,05881009 |
| 128 | 1 | 45675053 | 0,05766395 | 132 | 1 | 46065083 | 0,05797095 |
| 157 | 1 | 55173916 | 0,05766097 | 5933 | 6 | 424307850 | 0,05796349 |
| 1741 | 2 | 491400690 | 0,05733446 | 3658 | 4 | 246541697 | 0,05748568 |
| 3301 | 4 | 242019500 | 0,05724051 | 4439 | 5 | 292373890 | 0,05735926 |
| 2003 | 3 | 1882706 | 0,05696765 | 1024 | 2 | 4177434 | 0,05716170 |

marca grafada - associação com gene simulado

Fonte: Da autora (2019)

Tabela 12 – Bayes A × Bayes A modificado
cenário Poligênico , $h^2 = 0,5$

| Informações dos SNPS com efeito médio dos marcadores obtidos pelo método Bayes A pacote BGLR | | | | Informações dos SNPS com efeito médio dos marcadores obtidos pelo método Bayes A modificado | | | |
|--|------------|-----------|-----------------------|---|------------|-----------|-----------------------|
| Marca | Cromossomo | Posição | Efeito médio marcador | Marca | Cromossomo | Posição | Efeito médio marcador |
| 5826 | 6 | 386200335 | 2,32204403 | 5826 | 6 | 386200335 | 2,87187314 |
| 3480 | 4 | 245802107 | 0,67955884 | 1562 | 2 | 363489782 | 2,66348227 |
| 1562 | 2 | 363489782 | 0,21374070 | 3480 | 4 | 245802107 | 2,33180458 |
| 3261 | 4 | 212893509 | 0,08856406 | 4317 | 5 | 204516414 | 0,71706743 |
| 2749 | 3 | 596058675 | 0,08289701 | 3261 | 4 | 212893509 | 0,67441081 |
| 1646 | 2 | 423515013 | 0,07925501 | 2517 | 3 | 380243227 | 0,52410309 |
| 5846 | 6 | 393323235 | 0,07026384 | 1093 | 2 | 20954874 | 0,47881228 |
| 1358 | 2 | 214423504 | 0,06921913 | 3517 | 4 | 245955842 | 0,42582384 |
| 3960 | 4 | 247796507 | 0,06904547 | 5673 | 6 | 331710150 | 0,17162369 |
| 2517 | 3 | 380243227 | 0,06891925 | 4629 | 5 | 468337830 | 0,16197715 |
| 514 | 1 | 293632633 | 0,06641292 | 4337 | 5 | 209498142 | 0,15895269 |
| 4317 | 5 | 204516414 | 0,06553540 | 3971 | 4 | 247842212 | 0,13993051 |
| 1354 | 2 | 214263400 | 0,06541940 | 499 | 1 | 291031959 | 0,13129727 |
| 4554 | 5 | 398878380 | 0,06187091 | 4913 | 5 | 731357614 | 0,11305454 |
| 3517 | 4 | 245955842 | 0,06165870 | 4987 | 5 | 799890938 | 0,10782439 |
| 5577 | 6 | 297520230 | 0,06164649 | 5461 | 6 | 256207410 | 0,08583934 |
| 4265 | 5 | 184588958 | 0,06117547 | 1417 | 2 | 259874800 | 0,08581574 |
| 945 | 1 | 614932841 | 0,06067924 | 945 | 1 | 614932841 | 0,08291051 |
| 187 | 1 | 83595286 | 0,06066246 | 1295 | 2 | 186512163 | 0,08269730 |
| 499 | 1 | 291031959 | 0,06008913 | 5577 | 6 | 297520230 | 0,08082687 |
| 1210 | 2 | 128180134 | 0,05974159 | 4549 | 5 | 394247750 | 0,08029389 |
| 1338 | 2 | 206974778 | 0,05968785 | 3005 | 4 | 2637306 | 0,07721676 |
| 821 | 1 | 531839692 | 0,05965700 | 5277 | 6 | 190676730 | 0,07688821 |
| 1185 | 2 | 58354695 | 0,05736590 | 1263 | 2 | 169265278 | 0,07601629 |
| 1865 | 2 | 580009364 | 0,05719403 | 750 | 1 | 484262163 | 0,06941581 |
| 1313 | 2 | 195555350 | 0,05695609 | 3960 | 4 | 247796507 | 0,06880663 |
| 1417 | 2 | 259874800 | 0,05681460 | 5365 | 6 | 222017490 | 0,06301519 |
| 4913 | 5 | 731357614 | 0,05531022 | 162 | 1 | 56885616 | 0,06000659 |
| 3389 | 4 | 245424002 | 0,05493180 | 187 | 1 | 83595286 | 0,05977441 |
| 3864 | 4 | 247397627 | 0,05342407 | 2346 | 3 | 223513366 | 0,05967295 |
| 1093 | 2 | 20954874 | 0,05328539 | 4453 | 5 | 305339654 | 0,05963187 |
| 4138 | 5 | 62622237 | 0,05321744 | 4649 | 5 | 486860350 | 0,05947403 |
| 1213 | 2 | 128574019 | 0,05296734 | 2749 | 3 | 596058675 | 0,05824088 |
| 4453 | 5 | 305339654 | 0,05276296 | 898 | 1 | 583437857 | 0,05638806 |
| 5805 | 6 | 378721290 | 0,05238692 | 514 | 1 | 293632633 | 0,05614993 |
| 1955 | 2 | 644322111 | 0,05170282 | 1354 | 2 | 214263400 | 0,05591783 |
| 2936 | 3 | 770013368 | 0,05135616 | 1252 | 2 | 164561951 | 0,05489297 |
| 4648 | 5 | 485934224 | 0,05124320 | 1534 | 2 | 343481372 | 0,05352032 |
| 4649 | 5 | 486860350 | 0,05098012 | 1095 | 2 | 20958307 | 0,05170281 |
| 1278 | 2 | 173848369 | 0,05074562 | 4170 | 5 | 75234294 | 0,05099288 |
| 5862 | 6 | 399021555 | 0,05072464 | 2365 | 3 | 238846899 | 0,04707361 |
| 1252 | 2 | 164561951 | 0,05046685 | 4205 | 5 | 145121333 | 0,04662281 |
| 1550 | 2 | 354914749 | 0,05009883 | 1319 | 2 | 198756269 | 0,04658135 |
| 5277 | 6 | 190676730 | 0,04999836 | 821 | 1 | 531839692 | 0,04575879 |
| 4266 | 5 | 185596985 | 0,04985531 | 1243 | 2 | 159619959 | 0,04539647 |
| 1393 | 2 | 234574991 | 0,04959574 | 2936 | 3 | 770013368 | 0,04523382 |
| 5141 | 6 | 135105656 | 0,04933374 | 4266 | 5 | 185596985 | 0,04495615 |
| 530 | 1 | 295771335 | 0,04923345 | 3389 | 4 | 245424002 | 0,04495439 |
| 4987 | 5 | 799890938 | 0,04900788 | 4969 | 5 | 783220670 | 0,04392116 |
| 1881 | 2 | 591442741 | 0,04899076 | 4497 | 5 | 346089198 | 0,04390586 |
| 2072 | 3 | 53549251 | 0,04885925 | 1230 | 2 | 144528591 | 0,04343868 |
| 5673 | 6 | 331710150 | 0,04873467 | 4540 | 5 | 385912616 | 0,04237536 |
| 2877 | 3 | 715129267 | 0,04855417 | 1210 | 2 | 128180134 | 0,04158710 |
| 2088 | 3 | 69892547 | 0,04851102 | 2746 | 3 | 593267958 | 0,04108316 |
| 1991 | 2 | 670047210 | 0,04836910 | 4963 | 5 | 777663914 | 0,04006051 |
| 182 | 1 | 74579886 | 0,04805979 | 1358 | 2 | 214423504 | 0,03942285 |
| 2729 | 3 | 577453895 | 0,04747549 | 427 | 1 | 268373293 | 0,03914097 |
| 1653 | 2 | 428517115 | 0,04745500 | 3170 | 4 | 153293452 | 0,03813670 |
| 3773 | 4 | 247019522 | 0,04730621 | 996 | 1 | 649108249 | 0,03794374 |
| 1534 | 2 | 343481372 | 0,04730422 | 3773 | 4 | 247019522 | 0,03794044 |
| 1323 | 2 | 198757777 | 0,04730207 | 5962 | 6 | 434636055 | 0,03766265 |
| 5883 | 6 | 406500600 | 0,04715780 | 3918 | 4 | 247621997 | 0,03732103 |

marca grafada - associação com gene simulado

Fonte: Da autora (2019)

Tabela 13 – Bayes A × Bayes A modificado
cenário Poligênico, $h^2 = 0,8$

| Informações dos SNPS com efeito médio dos marcadores obtidos pelo método Bayes A pacote BGLR | | | | Informações dos SNPS com efeito médio dos marcadores obtidos pelo método Bayes A modificado | | | |
|--|------------|-----------|-----------------------|---|------------|-----------|-----------------------|
| Marca | Cromossomo | Posição | Efeito médio marcador | Marca | Cromossomo | Posição | Efeito médio marcador |
| 5826 | 6 | 386200335 | 2,39325846 | 5826 | 6 | 386200335 | 2,69335971 |
| 3480 | 4 | 245802107 | 1,90622709 | 1562 | 2 | 363489782 | 2,57673378 |
| 1562 | 2 | 363489782 | 1,66490364 | 3480 | 4 | 245802107 | 2,18827444 |
| 5033 | 6 | 75187712 | 0,14460003 | 2517 | 3 | 380243227 | 1,28672941 |
| 3643 | 4 | 246479372 | 0,12717845 | 4549 | 5 | 394247750 | 0,79155725 |
| 1093 | 2 | 20954874 | 0,12265745 | 1093 | 2 | 20954874 | 0,77649438 |
| 2517 | 3 | 380243227 | 0,10897191 | 3960 | 4 | 247796507 | 0,64174344 |
| 3517 | 4 | 245955842 | 0,09979534 | 3517 | 4 | 245955842 | 0,64008101 |
| 4629 | 5 | 468337830 | 0,09360603 | 3261 | 4 | 212893509 | 0,47672259 |
| 4549 | 5 | 394247750 | 0,09305793 | 3288 | 4 | 235779015 | 0,39265407 |
| 3960 | 4 | 247796507 | 0,08889688 | 4629 | 5 | 468337830 | 0,32945640 |
| 5560 | 6 | 291465765 | 0,08262290 | 5862 | 6 | 399021555 | 0,23058901 |
| 5862 | 6 | 399021555 | 0,07920124 | 5033 | 6 | 75187712 | 0,19459069 |
| 5290 | 6 | 195306615 | 0,07761892 | 1528 | 2 | 339193855 | 0,17697995 |
| 1646 | 2 | 423515013 | 0,07565332 | 5560 | 6 | 291465765 | 0,14548658 |
| 5636 | 6 | 318532785 | 0,07493743 | 5636 | 6 | 318532785 | 0,12447969 |
| 5612 | 6 | 309985305 | 0,07473683 | 382 | 1 | 262715518 | 0,10454140 |
| 2273 | 3 | 193565736 | 0,07428966 | 3476 | 4 | 245785487 | 0,10345652 |
| 4394 | 5 | 250698220 | 0,07402124 | 2776 | 3 | 621175128 | 0,09899893 |
| 89 | 1 | 28084996 | 0,07191011 | 3632 | 4 | 246433667 | 0,09362804 |
| 4459 | 5 | 310896410 | 0,07154023 | 4895 | 5 | 714687346 | 0,09027174 |
| 3261 | 4 | 212893509 | 0,06911439 | 4085 | 5 | 20200987 | 0,08985427 |
| 859 | 1 | 557303721 | 0,06883284 | 3032 | 4 | 28192994 | 0,07956836 |
| 1902 | 2 | 606449049 | 0,06859964 | 5590 | 6 | 302150115 | 0,07901568 |
| 1249 | 2 | 161879250 | 0,06854556 | 3406 | 4 | 245494637 | 0,07331959 |
| 5590 | 6 | 302150115 | 0,06829089 | 5643 | 6 | 321025800 | 0,07273018 |
| 966 | 1 | 629005068 | 0,06754910 | 5131 | 6 | 129927781 | 0,06780978 |
| 1880 | 2 | 590728155 | 0,06508941 | 4341 | 5 | 210890745 | 0,06511182 |
| 2289 | 3 | 200142461 | 0,06468038 | 3170 | 4 | 153293452 | 0,06451493 |
| 3476 | 4 | 245785487 | 0,06442367 | 2736 | 3 | 583965568 | 0,06306461 |
| 1087 | 2 | 20042233 | 0,06373584 | 3918 | 4 | 247621997 | 0,06256143 |
| 3918 | 4 | 247621997 | 0,06372235 | 89 | 1 | 28084996 | 0,05916935 |
| 4913 | 5 | 731357614 | 0,06333884 | 4849 | 5 | 672085550 | 0,05905166 |
| 3638 | 4 | 246458597 | 0,06305911 | 5416 | 6 | 240180885 | 0,05757233 |
| 3366 | 4 | 245328437 | 0,06295598 | 5109 | 6 | 113385418 | 0,05603946 |
| 187 | 1 | 83595286 | 0,06155839 | 5472 | 6 | 260125005 | 0,05546026 |
| 1153 | 2 | 39041117 | 0,06060107 | 3389 | 4 | 245424002 | 0,05368984 |
| 5461 | 6 | 256207410 | 0,06049723 | 859 | 1 | 557303721 | 0,05328594 |
| 3755 | 4 | 246944732 | 0,06033869 | 4459 | 5 | 310896410 | 0,05307614 |
| 2776 | 3 | 621175128 | 0,06032803 | 522 | 1 | 294646235 | 0,04997737 |
| 5145 | 6 | 137326250 | 0,05995661 | 5518 | 6 | 276507675 | 0,04931091 |
| 4593 | 5 | 434997294 | 0,05985198 | 3160 | 4 | 148408488 | 0,04828188 |
| 2608 | 3 | 464894976 | 0,05838805 | 3363 | 4 | 245315972 | 0,04776093 |
| 382 | 1 | 262715518 | 0,05800257 | 4648 | 5 | 485934224 | 0,04651553 |
| 2292 | 3 | 200201258 | 0,05794763 | 4650 | 5 | 487786476 | 0,04610854 |
| 3616 | 4 | 246367187 | 0,05701765 | 1902 | 2 | 606449049 | 0,04561702 |
| 4138 | 5 | 62622237 | 0,05686483 | 1646 | 2 | 423515013 | 0,04552106 |
| 3658 | 4 | 246541697 | 0,05671703 | 3318 | 4 | 245131700 | 0,04541531 |
| 2901 | 3 | 737455003 | 0,05617384 | 2279 | 3 | 194643731 | 0,04422672 |
| 322 | 1 | 223466873 | 0,05589089 | 4698 | 5 | 532240524 | 0,04421322 |
| 3288 | 4 | 235779015 | 0,05496678 | 4940 | 5 | 756363016 | 0,04321721 |
| 1759 | 2 | 504263240 | 0,05484391 | 5145 | 6 | 137326250 | 0,04299706 |
| 225 | 1 | 148694799 | 0,05460153 | 3755 | 4 | 246944732 | 0,04284377 |
| 3220 | 4 | 180309373 | 0,05458027 | 5206 | 6 | 164470369 | 0,04270003 |
| 3835 | 4 | 247277132 | 0,05452571 | 3448 | 4 | 245669147 | 0,04243763 |
| 4940 | 5 | 756363016 | 0,05395930 | 3845 | 4 | 247318682 | 0,04177470 |
| 5416 | 6 | 240180885 | 0,05370995 | 3382 | 4 | 245394917 | 0,04155739 |
| 771 | 1 | 498334390 | 0,05361059 | 3220 | 4 | 180309373 | 0,04131114 |
| 5805 | 6 | 378721290 | 0,05355605 | 5829 | 6 | 387268770 | 0,04087117 |
| 2998 | 3 | 827688186 | 0,05351080 | 105 | 1 | 35621803 | 0,03987407 |
| 728 | 1 | 469519830 | 0,05346140 | 3868 | 4 | 247414247 | 0,03943367 |
| 3295 | 4 | 238553534 | 0,05345099 | 3488 | 4 | 245835347 | 0,03860484 |

marca grifada - associação com gene simulado

Fonte: Da autora (2019)

Tabelas Bayes B x Bayes B modificado

Tabela 14 – Bayes B × Bayes B modificado
cenário Oligogênico, $h^2 = 0,2$

| Informações dos SNPS com efeito médio dos marcadores obtidos pelo método Bayes B pacote BGLR | | | | Informações dos SNPS com efeito médio dos marcadores obtidos pelo método Bayes B modificado | | | |
|--|------------|-----------|-----------------------|---|------------|-----------|-----------------------|
| Marca | Cromossomo | Posição | Efeito médio marcador | Marca | Cromossomo | Posição | Efeito médio marcador |
| 4004 | 5 | 2123914 | 0,98189606 | 4004 | 5 | 2123914 | 1,46652533 |
| 2983 | 3 | 813734601 | 0,06267106 | 2983 | 3 | 813734601 | 0,83467097 |
| 2897 | 3 | 733734047 | 0,03441984 | 1826 | 2 | 552140507 | 0,19444451 |
| 2896 | 3 | 732803808 | 0,03355121 | 4877 | 5 | 698017078 | 0,11714765 |
| 1705 | 2 | 465675591 | 0,03056261 | 1029 | 2 | 4679573 | 0,09168060 |
| 117 | 1 | 41505211 | 0,03037871 | 279 | 1 | 202620610 | 0,07782993 |
| 5285 | 6 | 193525890 | 0,02984817 | 1051 | 2 | 7482709 | 0,07262783 |
| 279 | 1 | 202620610 | 0,02700310 | 2256 | 3 | 186580149 | 0,07232988 |
| 1098 | 2 | 20958616 | 0,02687479 | 4045 | 5 | 9279776 | 0,06891233 |
| 2256 | 3 | 186580149 | 0,02586745 | 5259 | 6 | 184266120 | 0,06486697 |

marca grifada - associação com gene simulado

Fonte: Da autora (2019)

Tabela 15 – Bayes B × Bayes B modificado
cenário Oligogênico, $h^2 = 0,5$

| Informações dos SNPS com efeito médio dos marcadores obtidos pelo método Bayes B pacote BGLR | | | | Informações dos SNPS com efeito médio dos marcadores obtidos pelo método Bayes B modificado | | | |
|--|------------|-----------|-----------------------|---|------------|-----------|-----------------------|
| Marca | Cromossomo | Posição | Efeito médio marcador | Marca | Cromossomo | Posição | Efeito médio marcador |
| 4004 | 5 | 2123914 | 1,106274091 | 4004 | 5 | 2123914 | 1,187464568 |
| 3677 | 4 | 246620642 | 0,981411633 | 3677 | 4 | 246620642 | 1,080186587 |
| 4877 | 5 | 698017078 | 0,912558210 | 4877 | 5 | 698017078 | 1,072816052 |
| 193 | 1 | 88380847 | 0,042497960 | 193 | 1 | 88380847 | 0,118664425 |
| 833 | 1 | 539880964 | 0,022189562 | 713 | 1 | 459468239 | 0,052862673 |
| 3810 | 4 | 247173257 | 0,021229381 | 538 | 1 | 298413091 | 0,052742537 |
| 437 | 1 | 275982031 | 0,018955295 | 532 | 1 | 297315711 | 0,041384767 |
| 1830 | 2 | 554998851 | 0,016147924 | 547 | 1 | 348230637 | 0,037638761 |
| 577 | 1 | 368333818 | 0,016041541 | 577 | 1 | 368333818 | 0,035421382 |
| 3426 | 4 | 245577737 | 0,016037512 | 186 | 1 | 82563962 | 0,034793311 |

marca grifada - associação com gene simulado

Fonte: Da autora (2019)

Tabela 16 – Bayes B × Bayes B modificado
cenário Oligogênico, $h^2 = 0,8$

| Informações dos SNPS com efeito médio dos marcadores obtidos pelo método Bayes B pacote BGLR | | | | Informações dos SNPS com efeito médio dos marcadores obtidos pelo método Bayes B modificado | | | |
|--|------------|-----------|-----------------------|---|------------|-----------|-----------------------|
| Marca | Cromossomo | Posição | Efeito médio marcador | Marca | Cromossomo | Posição | Efeito médio marcador |
| 4004 | 5 | 2123914 | 1,407632625 | 4004 | 5 | 2123914 | 1,420697792 |
| 4877 | 5 | 698017078 | 1,050584570 | 4877 | 5 | 698017078 | 1,083414224 |
| 3677 | 4 | 246620642 | 0,698468467 | 3677 | 4 | 246620642 | 0,720254167 |
| 538 | 1 | 298413091 | 0,191981104 | 538 | 1 | 298413091 | 0,320050561 |
| 2088 | 3 | 69892547 | 0,165669179 | 2088 | 3 | 69892547 | 0,294117659 |
| 2000 | 2 | 676478485 | 0,018594832 | 2000 | 2 | 676478485 | 0,036536798 |
| 1137 | 2 | 32949215 | 0,015439857 | 4927 | 5 | 744323378 | 0,016666458 |
| 246 | 1 | 179852076 | 0,014893416 | 900 | 1 | 584778069 | 0,015813240 |
| 1516 | 2 | 330618822 | 0,013627881 | 1799 | 2 | 532846683 | 0,012853329 |
| 900 | 1 | 584778069 | 0,013060736 | 3373 | 4 | 245357522 | 0,012627843 |

marca grifada - associação com gene simulado

Fonte: Da autora (2019)

Tabela 17 – Bayes B × Bayes B modificado
cenário Poligênico, $h^2 = 0,2$

| Informações dos SNPS com efeito médio dos marcadores obtidos pelo método Bayes B pacote BGLR | | | | Informações dos SNPS com efeito médio dos marcadores obtidos pelo método Bayes B modificado | | | |
|--|------------|-----------|-----------------------|---|------------|-----------|-----------------------|
| Marca | Cromossomo | Posição | Efeito médio marcador | Marca | Cromossomo | Posição | Efeito médio marcador |
| 3480 | 4 | 245802107 | 0,14658296 | 3480 | 4 | 245802107 | 1,20226582 |
| 5826 | 6 | 386200335 | 0,09319739 | 5826 | 6 | 386200335 | 0,75210529 |
| 1408 | 2 | 253443525 | 0,08805326 | 3941 | 4 | 247717562 | 0,56963032 |
| 4966 | 5 | 780442292 | 0,08691759 | 5862 | 6 | 399021555 | 0,45244309 |
| 5862 | 6 | 399021555 | 0,07704084 | 5239 | 6 | 177143220 | 0,38969489 |
| 3766 | 4 | 246990437 | 0,07655428 | 4966 | 5 | 780442292 | 0,38491657 |
| 5801 | 6 | 377296710 | 0,07073901 | 3960 | 4 | 247796507 | 0,37574384 |
| 3718 | 4 | 246790997 | 0,06532876 | 5801 | 6 | 377296710 | 0,31756060 |
| 3941 | 4 | 247717562 | 0,06282612 | 3718 | 4 | 246790997 | 0,24561125 |
| 5693 | 6 | 338833050 | 0,05990207 | 1562 | 2 | 363489782 | 0,21549413 |
| 3810 | 4 | 247173257 | 0,05983004 | 2578 | 3 | 436987806 | 0,16907389 |
| 958 | 1 | 623644219 | 0,05930225 | 5990 | 6 | 444608115 | 0,15884425 |
| 3515 | 4 | 245947532 | 0,05898799 | 1573 | 2 | 371350229 | 0,15876128 |
| 3126 | 4 | 122795391 | 0,05700187 | 3810 | 4 | 247173257 | 0,13937819 |
| 5990 | 6 | 444608115 | 0,05667375 | 620 | 1 | 397148378 | 0,13311314 |
| 3517 | 4 | 245955842 | 0,05542110 | 2840 | 3 | 680710424 | 0,13059460 |
| 5763 | 6 | 363763200 | 0,05524020 | 805 | 1 | 521117995 | 0,12252440 |
| 3278 | 4 | 222331942 | 0,05521840 | 3399 | 4 | 245465552 | 0,11748785 |
| 4933 | 5 | 749880134 | 0,05487117 | 4365 | 5 | 223840566 | 0,11635857 |
| 522 | 1 | 294646235 | 0,05485302 | 3961 | 4 | 247800662 | 0,10481516 |
| 5269 | 6 | 187827570 | 0,05473278 | 5756 | 6 | 361270185 | 0,10186522 |
| 3104 | 4 | 90203655 | 0,05403883 | 2901 | 3 | 737455003 | 0,10154911 |
| 3833 | 4 | 247268822 | 0,05381275 | 3515 | 4 | 245947532 | 0,10109116 |
| 1573 | 2 | 371350229 | 0,05327615 | 5123 | 6 | 121839487 | 0,10021702 |
| 1935 | 2 | 630030390 | 0,05307141 | 3296 | 4 | 239237605 | 0,09299956 |
| 922 | 1 | 599520402 | 0,05220936 | 842 | 1 | 545911919 | 0,08968522 |
| 2101 | 3 | 110715956 | 0,05218883 | 3104 | 4 | 90203655 | 0,08943680 |
| 1965 | 2 | 651467972 | 0,05157461 | 1955 | 2 | 644322111 | 0,08840104 |
| 5920 | 6 | 419677965 | 0,05151767 | 3532 | 4 | 246018167 | 0,08695482 |
| 2570 | 3 | 429545894 | 0,05117624 | 656 | 1 | 421272195 | 0,08599731 |
| 1113 | 2 | 24893463 | 0,05112006 | 5350 | 6 | 216675315 | 0,08453391 |
| 5982 | 6 | 441758955 | 0,05110617 | 933 | 1 | 606891568 | 0,08105784 |
| 1203 | 2 | 89520665 | 0,05091627 | 5728 | 6 | 351298125 | 0,07360842 |
| 4439 | 5 | 292373890 | 0,05078089 | 4070 | 5 | 13981667 | 0,07263446 |
| 5234 | 6 | 175362495 | 0,05076658 | 5538 | 6 | 283630575 | 0,07111404 |
| 2143 | 3 | 144466243 | 0,05072018 | 3420 | 4 | 245552807 | 0,06817158 |
| 3420 | 4 | 245552807 | 0,05005037 | 3940 | 4 | 247713407 | 0,06656549 |
| 258 | 1 | 188048080 | 0,05003894 | 4085 | 5 | 20200987 | 0,06487036 |
| 5933 | 6 | 424307850 | 0,04998994 | 2998 | 3 | 827688186 | 0,06472023 |
| 4815 | 5 | 640597266 | 0,04963941 | 640 | 1 | 410550498 | 0,06376965 |
| 2840 | 3 | 680710424 | 0,04962548 | 1408 | 2 | 253443525 | 0,06325204 |
| 805 | 1 | 521117995 | 0,04958208 | 4960 | 5 | 774885536 | 0,06320560 |
| 2456 | 3 | 323498648 | 0,04942114 | 2726 | 3 | 574663178 | 0,06135130 |
| 4960 | 5 | 774885536 | 0,04931024 | 3680 | 4 | 246633107 | 0,06098197 |
| 5123 | 6 | 121839487 | 0,04916737 | 414 | 1 | 267615649 | 0,05922287 |
| 3780 | 4 | 247048607 | 0,04914249 | 794 | 1 | 513746829 | 0,05906761 |
| 3881 | 4 | 247468262 | 0,04904288 | 4411 | 5 | 266442362 | 0,05879761 |
| 2871 | 3 | 709547833 | 0,04898044 | 3964 | 4 | 247813127 | 0,05859215 |
| 3532 | 4 | 246018167 | 0,04876149 | 1776 | 2 | 516411203 | 0,05815227 |
| 1776 | 2 | 516411203 | 0,04869162 | 3766 | 4 | 246990437 | 0,05623558 |
| 4729 | 5 | 560950430 | 0,04869085 | 1402 | 2 | 249156009 | 0,05622039 |
| 2000 | 2 | 676478485 | 0,04850987 | 3044 | 4 | 35384247 | 0,05581642 |
| 382 | 1 | 262715518 | 0,04824713 | 3517 | 4 | 245955842 | 0,05496799 |
| 969 | 1 | 631015386 | 0,04777832 | 1774 | 2 | 514982031 | 0,05475462 |
| 5759 | 6 | 362338620 | 0,04768069 | 966 | 1 | 629005068 | 0,05354089 |
| 2913 | 3 | 748617871 | 0,04761570 | 2613 | 3 | 469546171 | 0,05320895 |
| 3658 | 4 | 246541697 | 0,04758027 | 697 | 1 | 448746543 | 0,05308645 |
| 3963 | 4 | 247808972 | 0,04757341 | 4231 | 5 | 169327292 | 0,05246492 |
| 5493 | 6 | 267604050 | 0,04752232 | 5183 | 6 | 156645441 | 0,05226587 |
| 5526 | 6 | 279356835 | 0,04745113 | 629 | 1 | 403179332 | 0,05181550 |

marca grifada - associação com gene simulado

marca grifada - associação com gene simulado

Fonte: Da autora (2019)

Tabela 18 – Bayes B × Bayes B modificado
cenário Poligênico , $h^2 = 0,5$

| Informações dos SNPS com efeito médio dos marcadores obtidos pelo método Bayes B pacote BGLR | | | | Informações dos SNPS com efeito médio dos marcadores obtidos pelo método Bayes B modificado | | | |
|--|------------|-----------|-----------------------|---|------------|-----------|-----------------------|
| Marca | Cromossomo | Posição | Efeito médio marcador | Marca | Cromossomo | Posição | Efeito médio marcador |
| 5826 | 6 | 386200335 | 2,39071999 | 5826 | 6 | 386200335 | 2,88199807 |
| 3480 | 4 | 245802107 | 1,16239964 | 1562 | 2 | 363489782 | 2,72893908 |
| 1562 | 2 | 363489782 | 1,01271911 | 3480 | 4 | 245802107 | 2,33768872 |
| 4317 | 5 | 204516414 | 0,26308285 | 3261 | 4 | 212893509 | 1,10085521 |
| 3261 | 4 | 212893509 | 0,25904899 | 4317 | 5 | 204516414 | 1,01291918 |
| 3517 | 4 | 245955842 | 0,22745188 | 2517 | 3 | 380243227 | 0,85900584 |
| 1646 | 2 | 423515013 | 0,15770354 | 1093 | 2 | 20954874 | 0,64858698 |
| 1093 | 2 | 20954874 | 0,11644133 | 3517 | 4 | 245955842 | 0,35926212 |
| 5461 | 6 | 256207410 | 0,11480515 | 4629 | 5 | 468337830 | 0,29681774 |
| 5846 | 6 | 393323235 | 0,11006904 | 514 | 1 | 293632633 | 0,21670450 |
| 2517 | 3 | 380243227 | 0,10895249 | 4913 | 5 | 731357614 | 0,18707007 |
| 1354 | 2 | 214263400 | 0,10167084 | 1417 | 2 | 259874800 | 0,14940769 |
| 4977 | 5 | 790629678 | 0,08512662 | 4337 | 5 | 209498142 | 0,14467985 |
| 1185 | 2 | 58354695 | 0,08263627 | 2346 | 3 | 223513366 | 0,11828482 |
| 4081 | 5 | 19973330 | 0,07932582 | 4987 | 5 | 799890938 | 0,11531797 |
| 945 | 1 | 614932841 | 0,07904553 | 187 | 1 | 83595286 | 0,10709324 |
| 5277 | 6 | 190676730 | 0,07585857 | 3773 | 4 | 247019522 | 0,09398777 |
| 1252 | 2 | 164561951 | 0,07367470 | 3971 | 4 | 247842212 | 0,09227150 |
| 5933 | 6 | 424307850 | 0,07318328 | 4549 | 5 | 394247750 | 0,09208964 |
| 1321 | 2 | 198757428 | 0,07279315 | 5577 | 6 | 297520230 | 0,09165486 |
| 1865 | 2 | 580009364 | 0,07177180 | 4341 | 5 | 210890745 | 0,08721330 |
| 677 | 1 | 435344422 | 0,07117438 | 1534 | 2 | 343481372 | 0,08431356 |
| 187 | 1 | 83595286 | 0,07017405 | 5673 | 6 | 331710150 | 0,08123734 |
| 499 | 1 | 291031959 | 0,07006015 | 5670 | 6 | 330641715 | 0,07783138 |
| 4629 | 5 | 468337830 | 0,06994469 | 1263 | 2 | 169265278 | 0,06961622 |
| 1191 | 2 | 62804122 | 0,06962805 | 2088 | 3 | 69892547 | 0,06637364 |
| 2749 | 3 | 596058675 | 0,06914042 | 5846 | 6 | 393323235 | 0,06403016 |
| 1338 | 2 | 206974778 | 0,06900470 | 5277 | 6 | 190676730 | 0,06160384 |
| 4266 | 5 | 185596985 | 0,06769650 | 4540 | 5 | 385912616 | 0,06040912 |
| 4369 | 5 | 227545070 | 0,06700470 | 1295 | 2 | 186512163 | 0,05752053 |
| 3389 | 4 | 245424002 | 0,06685639 | 3960 | 4 | 247796507 | 0,05732947 |
| 3971 | 4 | 247842212 | 0,06634282 | 4098 | 5 | 30239292 | 0,05568821 |
| 3972 | 4 | 247846367 | 0,06592994 | 1230 | 2 | 144528591 | 0,05392123 |
| 821 | 1 | 531839692 | 0,06490813 | 5884 | 6 | 406856745 | 0,05279349 |
| 4969 | 5 | 783220670 | 0,06287196 | 2946 | 3 | 779315758 | 0,05263086 |
| 2842 | 3 | 682570902 | 0,06256251 | 4138 | 5 | 62622237 | 0,05144616 |
| 427 | 1 | 268373293 | 0,06252005 | 1095 | 2 | 20958307 | 0,05107890 |
| 3736 | 4 | 246865787 | 0,06197892 | 1321 | 2 | 198757428 | 0,05053855 |
| 514 | 1 | 293632633 | 0,06015120 | 2800 | 3 | 643500864 | 0,04970587 |
| 1095 | 2 | 20958307 | 0,05969377 | 4648 | 5 | 485934224 | 0,04918979 |
| 1090 | 2 | 20722170 | 0,05935112 | 671 | 1 | 431323786 | 0,04630457 |
| 3976 | 4 | 247862987 | 0,05933717 | 5461 | 6 | 256207410 | 0,04599792 |
| 1980 | 2 | 662186763 | 0,05876494 | 5862 | 6 | 399021555 | 0,04478699 |
| 2493 | 3 | 357917491 | 0,05827585 | 1561 | 2 | 362775196 | 0,04468010 |
| 1404 | 2 | 250585181 | 0,05769171 | 5962 | 6 | 434636055 | 0,04364331 |
| 5884 | 6 | 406856745 | 0,05734149 | 4266 | 5 | 185596985 | 0,04200111 |
| 5131 | 6 | 129927781 | 0,05693318 | 4081 | 5 | 19973330 | 0,04185745 |
| 555 | 1 | 353591485 | 0,05668442 | 1646 | 2 | 423515013 | 0,03983977 |
| 3133 | 4 | 128632208 | 0,05651398 | 970 | 1 | 631685492 | 0,03840169 |
| 5235 | 6 | 175718640 | 0,05647650 | 43 | 1 | 10067746 | 0,03811319 |
| 5815 | 6 | 382282740 | 0,05640626 | 5050 | 6 | 80877074 | 0,03556321 |
| 4219 | 5 | 166468431 | 0,05587958 | 4838 | 5 | 661898164 | 0,03521571 |
| 2877 | 3 | 715129267 | 0,05544448 | 4205 | 5 | 145121333 | 0,03492502 |
| 1143 | 2 | 37328160 | 0,05493753 | 100 | 1 | 34479054 | 0,03488546 |
| 1203 | 2 | 89520665 | 0,05492586 | 4670 | 5 | 506308996 | 0,03446493 |
| 3633 | 4 | 246437822 | 0,05482289 | 4170 | 5 | 75234294 | 0,03359091 |
| 182 | 1 | 74579886 | 0,05479871 | 2210 | 3 | 171935683 | 0,03357751 |
| 162 | 1 | 56885616 | 0,05476782 | 1252 | 2 | 164561951 | 0,03301484 |
| 1140 | 2 | 35753260 | 0,05466744 | 1167 | 2 | 44606542 | 0,03295700 |
| 5724 | 6 | 349873545 | 0,05459916 | 2274 | 3 | 193566873 | 0,03257957 |

marca grifada - associação com gene simulado

Fonte: Da autora (2019)

Tabela 19 – Bayes B × Bayes B modificado
cenário Poligênico, $h^2 = 0,8$

| Informações dos SNPS com efeito médio dos marcadores obtidos pelo método Bayes B pacote BGLR | | | | Informações dos SNPS com efeito médio dos marcadores obtidos pelo método Bayes B modificado | | | |
|--|------------|-----------|-----------------------|---|------------|-----------|-----------------------|
| Marca | Cromossomo | Posição | Efeito médio marcador | Marca | Cromossomo | Posição | Efeito médio marcador |
| 5826 | 6 | 386200335 | 2,31209141 | 5826 | 6 | 386200335 | 2,78452126 |
| 3480 | 4 | 245802107 | 2,03291471 | 1562 | 2 | 363489782 | 2,53628860 |
| 1562 | 2 | 363489782 | 1,77746691 | 3480 | 4 | 245802107 | 2,28691431 |
| 2517 | 3 | 380243227 | 0,61615945 | 2517 | 3 | 380243227 | 1,46402879 |
| 5033 | 6 | 75187712 | 0,32523240 | 3261 | 4 | 212893509 | 1,39874094 |
| 4629 | 5 | 468337830 | 0,28088000 | 4549 | 5 | 394247750 | 1,26670823 |
| 1093 | 2 | 20954874 | 0,25416947 | 4629 | 5 | 468337830 | 1,12502322 |
| 3643 | 4 | 246479372 | 0,17629336 | 1093 | 2 | 20954874 | 1,03940426 |
| 5862 | 6 | 399021555 | 0,17150306 | 3960 | 4 | 247796507 | 1,01058824 |
| 3517 | 4 | 245955842 | 0,15636298 | 3288 | 4 | 235779015 | 0,60098682 |
| 4549 | 5 | 394247750 | 0,14118362 | 5560 | 6 | 291465765 | 0,30187657 |
| 3288 | 4 | 235779015 | 0,13605683 | 2279 | 3 | 194643731 | 0,27333640 |
| 3261 | 4 | 212893509 | 0,12915640 | 2289 | 3 | 200142461 | 0,25767252 |
| 859 | 1 | 557303721 | 0,11549034 | 5033 | 6 | 75187712 | 0,23081905 |
| 3918 | 4 | 247621997 | 0,11285098 | 5862 | 6 | 399021555 | 0,21325630 |
| 89 | 1 | 28084996 | 0,10785534 | 3476 | 4 | 245785487 | 0,19756249 |
| 5590 | 6 | 302150115 | 0,09666319 | 3337 | 4 | 245207942 | 0,16889058 |
| 1686 | 2 | 452098456 | 0,09619786 | 4648 | 5 | 485934224 | 0,16488947 |
| 1646 | 2 | 423515013 | 0,08894054 | 5643 | 6 | 321025800 | 0,15115927 |
| 3960 | 4 | 247796507 | 0,08886736 | 3318 | 4 | 245131700 | 0,12351988 |
| 4394 | 5 | 250698220 | 0,08797212 | 1528 | 2 | 339193855 | 0,11018048 |
| 1249 | 2 | 161879250 | 0,08567574 | 5590 | 6 | 302150115 | 0,10539827 |
| 382 | 1 | 262715518 | 0,08523822 | 1225 | 2 | 140264872 | 0,10426519 |
| 1153 | 2 | 39041117 | 0,08463196 | 3252 | 4 | 206492925 | 0,10024888 |
| 5560 | 6 | 291465765 | 0,08364097 | 5109 | 6 | 113385418 | 0,09966367 |
| 966 | 1 | 629005068 | 0,08354173 | 5472 | 6 | 260125005 | 0,09525918 |
| 3476 | 4 | 245785487 | 0,08216741 | 4085 | 5 | 20200987 | 0,09442426 |
| 4940 | 5 | 756363016 | 0,08171415 | 2736 | 3 | 583965568 | 0,09318871 |
| 158 | 1 | 56175806 | 0,08164360 | 382 | 1 | 262715518 | 0,07854643 |
| 1087 | 2 | 20042233 | 0,07932709 | 3488 | 4 | 245835347 | 0,07292818 |
| 3538 | 4 | 246043097 | 0,07783412 | 5636 | 6 | 318532785 | 0,07225943 |
| 4566 | 5 | 409991892 | 0,07776864 | 1087 | 2 | 20042233 | 0,06580713 |
| 1552 | 2 | 356343921 | 0,07712431 | 4459 | 5 | 310896410 | 0,06378011 |
| 2261 | 3 | 187604478 | 0,07675882 | 4951 | 5 | 766550402 | 0,06181961 |
| 5416 | 6 | 240180885 | 0,07667240 | 3755 | 4 | 246944732 | 0,06051358 |
| 105 | 1 | 35621803 | 0,07551088 | 3406 | 4 | 245494637 | 0,05841696 |
| 3616 | 4 | 246367187 | 0,07480834 | 4895 | 5 | 714687346 | 0,05788273 |
| 2273 | 3 | 193565736 | 0,07460160 | 4820 | 5 | 645227896 | 0,05764271 |
| 5465 | 6 | 257631990 | 0,07379959 | 5518 | 6 | 276507675 | 0,05732786 |
| 5131 | 6 | 129927781 | 0,07250235 | 4849 | 5 | 672085550 | 0,05670412 |
| 4951 | 5 | 766550402 | 0,07131607 | 3682 | 4 | 246641417 | 0,05614454 |
| 3632 | 4 | 246433667 | 0,06969893 | 251 | 1 | 183647544 | 0,05289388 |
| 5882 | 6 | 406144455 | 0,06959228 | 3845 | 4 | 247318682 | 0,05192228 |
| 199 | 1 | 90772644 | 0,06849733 | 89 | 1 | 28084996 | 0,05051600 |
| 5145 | 6 | 137326250 | 0,06814621 | 3350 | 4 | 245261957 | 0,05016864 |
| 2292 | 3 | 200201258 | 0,06757530 | 5417 | 6 | 240537030 | 0,04957675 |
| 4820 | 5 | 645227896 | 0,06719873 | 1646 | 2 | 423515013 | 0,04909307 |
| 1964 | 2 | 650753386 | 0,06695361 | 3679 | 4 | 246628952 | 0,04870569 |
| 1228 | 2 | 144527921 | 0,06526687 | 3205 | 4 | 173687725 | 0,04755869 |
| 2086 | 3 | 69079684 | 0,06414500 | 3032 | 4 | 28192994 | 0,04486354 |
| 3218 | 4 | 178735388 | 0,06337462 | 5461 | 6 | 256207410 | 0,04146100 |
| 5636 | 6 | 318532785 | 0,06279653 | 159 | 1 | 56846640 | 0,04043599 |
| 3318 | 4 | 245131700 | 0,06262185 | 4266 | 5 | 185596985 | 0,03976024 |
| 1865 | 2 | 580009364 | 0,06242233 | 522 | 1 | 294646235 | 0,03752937 |
| 5472 | 6 | 260125005 | 0,06239465 | 5416 | 6 | 240180885 | 0,03668346 |
| 3389 | 4 | 245424002 | 0,06236930 | 3898 | 4 | 247538897 | 0,03619043 |
| 486 | 1 | 286642725 | 0,06185754 | 233 | 1 | 173860693 | 0,03473378 |
| 728 | 1 | 469519830 | 0,06162497 | 4853 | 5 | 675790054 | 0,03363149 |
| 1846 | 2 | 566432229 | 0,06018398 | 1249 | 2 | 161879250 | 0,03313616 |
| 2289 | 3 | 200142461 | 0,05977891 | 886 | 1 | 575396584 | 0,03284292 |

marca grifada - associação com gene simulado

Fonte: Da autora (2019)

Tabelas BSSV x BSSV modificado

Tabela 20 – BSSV × BSSV modificado
cenário Oligogênico, $h^2 = 0,2$

| Informações dos SNPS com efeito médio dos marcadores obtidos pelo método BSSV | | | | Informações dos SNPS com efeito médio dos marcadores obtidos pelo método BSSV modificado | | | |
|---|------------|-----------|-----------------------|--|------------|-----------|-----------------------|
| Marca | Cromossomo | Posição | Efeito médio marcador | Marca | Cromossomo | Posição | Efeito médio marcador |
| 4004 | 5 | 2123914 | 1,215114244 | 4004 | 5 | 2123914 | 1,471020669 |
| 2983 | 3 | 813734601 | 0,433492862 | 2983 | 3 | 813734601 | 0,637994617 |
| 2897 | 3 | 733734047 | 0,141951270 | 2897 | 3 | 733734047 | 0,235071657 |
| 1029 | 2 | 4679573 | 0,067100930 | 1029 | 2 | 4679573 | 0,150197998 |
| 4877 | 5 | 698017078 | 0,044308722 | 279 | 1 | 202620610 | 0,096861852 |
| 4260 | 5 | 181672656 | 0,034296048 | 2896 | 3 | 732803808 | 0,093631589 |
| 3443 | 4 | 245648372 | 0,033598287 | 4877 | 5 | 698017078 | 0,076268172 |
| 2256 | 3 | 186580149 | 0,032559167 | 922 | 1 | 599520402 | 0,060835485 |
| 2896 | 3 | 732803808 | 0,031857981 | 4045 | 5 | 9279776 | 0,060778201 |
| 4621 | 5 | 460928822 | 0,022493705 | 4889 | 5 | 709130590 | 0,052278775 |

marca grifada - associação com gene simulado

Fonte: Da autora (2019)

Tabela 21 – BSSV × BSSV modificado
cenário Oligogênico, $h^2 = 0,5$

| Informações dos SNPS com efeito médio dos marcadores obtidos pelo método BSSV | | | | Informações dos SNPS com efeito médio dos marcadores obtidos pelo método BSSV modificado | | | |
|---|------------|-----------|-----------------------|--|------------|-----------|-----------------------|
| Marca | Cromossomo | Posição | Efeito médio marcador | Marca | Cromossomo | Posição | Efeito médio marcador |
| 4004 | 5 | 2123914 | 1,123826933 | 4004 | 5 | 2123914 | 1,191522241 |
| 3677 | 4 | 246620642 | 1,086561710 | 3677 | 4 | 246620642 | 1,091060350 |
| 4877 | 5 | 698017078 | 0,941017876 | 4877 | 5 | 698017078 | 1,082834735 |
| 538 | 1 | 298413091 | 0,009844054 | 193 | 1 | 88380847 | 0,203138196 |
| 435 | 1 | 273531917 | 0,009219474 | 1968 | 2 | 653611730 | 0,071879279 |
| 3792 | 4 | 247098467 | 0,009001584 | 713 | 1 | 459468239 | 0,045664457 |
| 212 | 1 | 102663772 | 0,008014600 | 532 | 1 | 297315711 | 0,044660272 |
| 410 | 1 | 266955761 | 0,007826519 | 877 | 1 | 569365630 | 0,042813521 |
| 186 | 1 | 82563962 | 0,007689784 | 538 | 1 | 298413091 | 0,033219901 |
| 193 | 1 | 88380847 | 0,006232948 | 5819 | 6 | 383707320 | 0,029790107 |

marca grifada - associação com gene simulado

Fonte: Da autora (2019)

Tabela 22 – BSSV × BSSV modificado
cenário Oligogênico, $h^2 = 0,8$

| Informações dos SNPS com efeito médio dos marcadores obtidos pelo método BSSV | | | | Informações dos SNPS com efeito médio dos marcadores obtidos pelo método BSSV modificado | | | |
|---|------------|-----------|-----------------------|--|------------|-----------|-----------------------|
| Marca | Cromossomo | Posição | Efeito médio marcador | Marca | Cromossomo | Posição | Efeito médio marcador |
| 4004 | 5 | 2123914 | 1,322893590 | 4004 | 5 | 2123914 | 1,417498032 |
| 4877 | 5 | 698017078 | 1,013647238 | 4877 | 5 | 698017078 | 1,087013940 |
| 3677 | 4 | 246620642 | 0,829483510 | 3677 | 4 | 246620642 | 0,724324078 |
| 538 | 1 | 298413091 | 0,022442881 | 2088 | 3 | 69892547 | 0,394194631 |
| 410 | 1 | 266955761 | 0,011630282 | 538 | 1 | 298413091 | 0,336789299 |
| 2000 | 2 | 676478485 | 0,005815726 | 2000 | 2 | 676478485 | 0,139605863 |
| 3777 | 4 | 247036142 | 0,005274398 | 2855 | 3 | 694664009 | 0,018343766 |
| 139 | 1 | 47280990 | 0,004853056 | 4927 | 5 | 744323378 | 0,017299150 |
| 9 | 1 | 3205262 | 0,004637922 | 5764 | 6 | 364119345 | 0,013764663 |
| 4925 | 5 | 742471126 | 0,004425646 | 900 | 1 | 584778069 | 0,011230743 |

marca grifada - associação com gene simulado

Fonte: Da autora (2019)

Tabela 23 – BSSV × BSSV modificado
cenário Poligênico, $h^2 = 0,2$

| Informações dos SNPS com efeito médio dos marcadores obtidos pelo BSSV | | | | Informações dos SNPS com efeito médio dos marcadores obtidos pelo BSSV modificado | | | |
|--|------------|-----------|-----------------------|---|------------|-----------|-----------------------|
| Marca | Cromossomo | Posição | Efeito médio marcador | Marca | Cromossomo | Posição | Efeito médio marcador |
| 5862 | 6 | 399021555 | 1,29345538 | 5826 | 6 | 386200335 | 1,03030309 |
| 5826 | 6 | 386200335 | 0,26185016 | 3480 | 4 | 245802107 | 0,99335121 |
| 3517 | 4 | 245955842 | 0,24631382 | 5862 | 6 | 399021555 | 0,41495582 |
| 3766 | 4 | 246990437 | 0,22272106 | 3517 | 4 | 245955842 | 0,34182990 |
| 5175 | 6 | 151958886 | 0,19869053 | 2840 | 3 | 680710424 | 0,32647015 |
| 3480 | 4 | 245802107 | 0,13785165 | 5728 | 6 | 351298125 | 0,30760986 |
| 5269 | 6 | 187827570 | 0,13235409 | 1408 | 2 | 253443525 | 0,28198065 |
| 5801 | 6 | 377296710 | 0,12419019 | 1562 | 2 | 363489782 | 0,27458680 |
| 805 | 1 | 521117995 | 0,09928076 | 5801 | 6 | 377296710 | 0,25263526 |
| 5728 | 6 | 351298125 | 0,09299846 | 3941 | 4 | 247717562 | 0,24030416 |
| 3810 | 4 | 247173257 | 0,09249162 | 4966 | 5 | 780442292 | 0,21968745 |
| 5239 | 6 | 177143220 | 0,07867484 | 5239 | 6 | 177143220 | 0,20507530 |
| 3399 | 4 | 245465552 | 0,06844750 | 620 | 1 | 397148378 | 0,17990112 |
| 2840 | 3 | 680710424 | 0,05582415 | 5269 | 6 | 187827570 | 0,17193802 |
| 5277 | 6 | 190676730 | 0,05331462 | 3658 | 4 | 246541697 | 0,17173028 |
| 2578 | 3 | 436987806 | 0,05284295 | 5123 | 6 | 121839487 | 0,16152352 |
| 2726 | 3 | 574663178 | 0,05195449 | 3718 | 4 | 246790997 | 0,15779983 |
| 5291 | 6 | 195662760 | 0,05130429 | 3766 | 4 | 246990437 | 0,13960862 |
| 4087 | 5 | 21770965 | 0,04839958 | 2578 | 3 | 436987806 | 0,12787366 |
| 2968 | 3 | 799781016 | 0,04811588 | 805 | 1 | 521117995 | 0,12709856 |
| 4838 | 5 | 661898164 | 0,04752063 | 3810 | 4 | 247173257 | 0,12412826 |
| 5756 | 6 | 361270185 | 0,04684513 | 842 | 1 | 545911919 | 0,12023593 |
| 1573 | 2 | 371350229 | 0,04647682 | 4087 | 5 | 21770965 | 0,11207576 |
| 3941 | 4 | 247717562 | 0,04623623 | 1268 | 2 | 169989408 | 0,10992083 |
| 5478 | 6 | 262261875 | 0,04520121 | 4058 | 5 | 11737498 | 0,10726694 |
| 3126 | 4 | 122795391 | 0,04270957 | 3296 | 4 | 239237605 | 0,10563962 |
| 3960 | 4 | 247796507 | 0,04035216 | 1406 | 2 | 252014353 | 0,09865264 |
| 5123 | 6 | 121839487 | 0,03912706 | 1573 | 2 | 371350229 | 0,09578311 |
| 5379 | 6 | 227003520 | 0,03872720 | 1657 | 2 | 431375459 | 0,09256241 |
| 3164 | 4 | 149588842 | 0,03630764 | 3515 | 4 | 245947532 | 0,08794077 |
| 1408 | 2 | 253443525 | 0,03460706 | 1941 | 2 | 634317906 | 0,08328377 |
| 37 | 1 | 9300541 | 0,03268534 | 5210 | 6 | 165724924 | 0,07567135 |
| 1406 | 2 | 252014353 | 0,03256397 | 3387 | 4 | 245415692 | 0,07379062 |
| 699 | 1 | 450086755 | 0,03147569 | 3960 | 4 | 247796507 | 0,07274150 |
| 4704 | 5 | 537797280 | 0,03117039 | 3399 | 4 | 245465552 | 0,07196399 |
| 842 | 1 | 545911919 | 0,02933925 | 2584 | 3 | 442569240 | 0,06988954 |
| 3420 | 4 | 245552807 | 0,02913208 | 382 | 1 | 262715518 | 0,06897437 |
| 2221 | 3 | 178234880 | 0,02907926 | 3846 | 4 | 247322837 | 0,06810345 |
| 5818 | 6 | 383351175 | 0,02885001 | 4380 | 5 | 237732456 | 0,06806182 |
| 933 | 1 | 606891568 | 0,02833495 | 2726 | 3 | 574663178 | 0,06706580 |
| 1562 | 2 | 363489782 | 0,02775874 | 5756 | 6 | 361270185 | 0,06663788 |
| 4928 | 5 | 745249504 | 0,02689853 | 4704 | 5 | 537797280 | 0,06531922 |
| 2393 | 3 | 264893591 | 0,02616913 | 3001 | 4 | 139753 | 0,06231611 |
| 620 | 1 | 397148378 | 0,02605647 | 2117 | 3 | 122632201 | 0,06207695 |
| 3658 | 4 | 246541697 | 0,02604357 | 3420 | 4 | 245552807 | 0,06038597 |
| 3437 | 4 | 245623442 | 0,02583229 | 3044 | 4 | 35384247 | 0,06018288 |
| 5693 | 6 | 338833050 | 0,02493277 | 5990 | 6 | 444608115 | 0,05899636 |
| 2613 | 3 | 469546171 | 0,02436947 | 1865 | 2 | 580009364 | 0,05862954 |
| 1268 | 2 | 169989408 | 0,02395064 | 2613 | 3 | 469546171 | 0,05797100 |
| 5990 | 6 | 444608115 | 0,02367736 | 3626 | 4 | 246408737 | 0,05750251 |
| 5357 | 6 | 219168330 | 0,02311281 | 5708 | 6 | 344175225 | 0,05696143 |
| 629 | 1 | 403179332 | 0,02307402 | 4070 | 5 | 13981667 | 0,05672457 |
| 2584 | 3 | 442569240 | 0,02236303 | 629 | 1 | 403179332 | 0,05468289 |
| 5708 | 6 | 344175225 | 0,02228276 | 5478 | 6 | 262261875 | 0,05422178 |
| 3736 | 4 | 246865787 | 0,02226718 | 5350 | 6 | 216675315 | 0,05195741 |
| 94 | 1 | 32583282 | 0,02203001 | 5693 | 6 | 338833050 | 0,05187646 |
| 3044 | 4 | 35384247 | 0,02194537 | 1776 | 2 | 516411203 | 0,05100499 |
| 177 | 1 | 64261162 | 0,02189480 | 2221 | 3 | 178234880 | 0,05087678 |
| 3532 | 4 | 246018167 | 0,02177482 | 37 | 1 | 9300541 | 0,05009267 |
| 5282 | 6 | 192457455 | 0,02155006 | 745 | 1 | 480911633 | 0,04997990 |

marca grifada - associação com gene simulado

Fonte: Da autora (2019)

Tabela 24 – BSSV × BSSV modificado
cenário Poligênico, $h^2 = 0,5$

| Informações dos SNPS com efeito médio dos marcadores obtidos pelo método BSSV | | | | Informações dos SNPS com efeito médio dos marcadores obtidos pelo método BSSV modificado | | | |
|---|------------|-----------|-----------------------|--|------------|-----------|-----------------------|
| Marca | Cromossomo | Posição | Efeito médio marcador | Marca | Cromossomo | Posição | Efeito médio marcador |
| 5826 | 6 | 386200335 | 2,475321760 | 5826 | 6 | 386200335 | 2,92711081 |
| 3480 | 4 | 245802107 | 0,707792891 | 1562 | 2 | 363489782 | 2,72128697 |
| 1562 | 2 | 363489782 | 0,695997677 | 3480 | 4 | 245802107 | 2,34732722 |
| 3517 | 4 | 245955842 | 0,629687042 | 4317 | 5 | 204516414 | 1,07202378 |
| 5673 | 6 | 331710150 | 0,553161940 | 3261 | 4 | 212893509 | 0,56695329 |
| 3261 | 4 | 212893509 | 0,281144091 | 1093 | 2 | 20954874 | 0,56039441 |
| 4629 | 5 | 468337830 | 0,195517804 | 3517 | 4 | 245955842 | 0,48284678 |
| 4337 | 5 | 209498142 | 0,116996705 | 2749 | 3 | 596058675 | 0,32205995 |
| 4977 | 5 | 790629678 | 0,100486328 | 2517 | 3 | 380243227 | 0,28255891 |
| 1865 | 2 | 580009364 | 0,070553405 | 1417 | 2 | 259874800 | 0,22203619 |
| 4895 | 5 | 714687346 | 0,070147228 | 5673 | 6 | 331710150 | 0,19921730 |
| 5277 | 6 | 190676730 | 0,064044154 | 3960 | 4 | 247796507 | 0,18634852 |
| 3301 | 4 | 242019500 | 0,047049592 | 3971 | 4 | 247842212 | 0,18056280 |
| 499 | 1 | 291031959 | 0,046746186 | 2346 | 3 | 223513366 | 0,17118696 |
| 3736 | 4 | 246865787 | 0,046643890 | 5461 | 6 | 256207410 | 0,13161507 |
| 4170 | 5 | 75234294 | 0,044593081 | 4987 | 5 | 799890938 | 0,11820071 |
| 1321 | 2 | 198757428 | 0,035718870 | 4913 | 5 | 731357614 | 0,11526312 |
| 4091 | 5 | 21771433 | 0,032953280 | 1252 | 2 | 164561951 | 0,11346914 |
| 2045 | 3 | 27199028 | 0,030844454 | 499 | 1 | 291031959 | 0,11309829 |
| 3096 | 4 | 76123039 | 0,030175921 | 4549 | 5 | 394247750 | 0,10432284 |
| 5986 | 6 | 443183535 | 0,029746229 | 5577 | 6 | 297520230 | 0,09772629 |
| 4649 | 5 | 486860350 | 0,028395482 | 187 | 1 | 83595286 | 0,09485042 |
| 5724 | 6 | 349873545 | 0,027960851 | 1095 | 2 | 20958307 | 0,09188900 |
| 3773 | 4 | 247019522 | 0,027724582 | 945 | 1 | 614932841 | 0,08676008 |
| 4317 | 5 | 204516414 | 0,027316668 | 2602 | 3 | 459313542 | 0,08378122 |
| 726 | 1 | 468179618 | 0,025853136 | 4138 | 5 | 62622237 | 0,07364580 |
| 2749 | 3 | 596058675 | 0,025418828 | 1646 | 2 | 423515013 | 0,06963642 |
| 3810 | 4 | 247173257 | 0,025325488 | 4629 | 5 | 468337830 | 0,06414368 |
| 3224 | 4 | 184373943 | 0,025055857 | 3315 | 4 | 245051599 | 0,06207592 |
| 1139 | 2 | 35040818 | 0,024851203 | 2088 | 3 | 69892547 | 0,06167634 |
| 4341 | 5 | 210890745 | 0,024329038 | 4081 | 5 | 19973330 | 0,06087298 |
| 5862 | 6 | 399021555 | 0,023879855 | 5670 | 6 | 330641715 | 0,06071885 |
| 4838 | 5 | 661898164 | 0,023740875 | 1534 | 2 | 343481372 | 0,05968492 |
| 3992 | 4 | 247929467 | 0,023703316 | 5724 | 6 | 349873545 | 0,05952437 |
| 4453 | 5 | 305339654 | 0,023371635 | 4648 | 5 | 485934224 | 0,05797836 |
| 1178 | 2 | 51785620 | 0,022791448 | 4554 | 5 | 398878380 | 0,05786321 |
| 1093 | 2 | 20954874 | 0,022693416 | 3773 | 4 | 247019522 | 0,05744555 |
| 3634 | 4 | 246441977 | 0,021534670 | 1230 | 2 | 144528591 | 0,05582280 |
| 4063 | 5 | 11992708 | 0,021478673 | 2245 | 3 | 185290309 | 0,05393972 |
| 1148 | 2 | 39040619 | 0,021125148 | 514 | 1 | 293632633 | 0,05251750 |
| 4290 | 5 | 193521948 | 0,020770884 | 1263 | 2 | 169265278 | 0,05219738 |
| 2517 | 3 | 380243227 | 0,020288587 | 4170 | 5 | 75234294 | 0,05179962 |
| 4554 | 5 | 398878380 | 0,019888518 | 2800 | 3 | 643500864 | 0,05005241 |
| 2090 | 3 | 82173052 | 0,019670927 | 2874 | 3 | 712338550 | 0,04842075 |
| 2406 | 3 | 276986698 | 0,019545773 | 4341 | 5 | 210890745 | 0,04809292 |
| 3608 | 4 | 246333947 | 0,019447337 | 1319 | 2 | 198756269 | 0,04798670 |
| 5931 | 6 | 423595560 | 0,018951235 | 4838 | 5 | 661898164 | 0,04763821 |
| 3864 | 4 | 247397627 | 0,018906729 | 5846 | 6 | 393323235 | 0,04702263 |
| 2877 | 3 | 715129267 | 0,018820164 | 4453 | 5 | 305339654 | 0,04602231 |
| 2840 | 3 | 680710424 | 0,018703346 | 2746 | 3 | 593267958 | 0,04561555 |
| 1646 | 2 | 423515013 | 0,018489272 | 1275 | 2 | 170899849 | 0,04539455 |
| 3170 | 4 | 153293452 | 0,018330572 | 5794 | 6 | 374803695 | 0,04328092 |
| 4913 | 5 | 731357614 | 0,018095228 | 4337 | 5 | 209498142 | 0,04324375 |
| 530 | 1 | 295771335 | 0,018063036 | 1295 | 2 | 186512163 | 0,04286134 |
| 5577 | 6 | 297520230 | 0,017858776 | 43 | 1 | 10067746 | 0,04245484 |
| 4721 | 5 | 553541422 | 0,017734497 | 970 | 1 | 631685492 | 0,04241325 |
| 354 | 1 | 240574310 | 0,017567317 | 5660 | 6 | 327080265 | 0,04202189 |
| 2072 | 3 | 53549251 | 0,017441999 | 1653 | 2 | 428517115 | 0,04197444 |
| 1506 | 2 | 323472961 | 0,017318943 | 1865 | 2 | 580009364 | 0,04163309 |
| 5846 | 6 | 393323235 | 0,017177582 | 1167 | 2 | 44606542 | 0,04148404 |

marca grifada - associação com gene simulado

Fonte: Da autora (2019)

Tabela 25 – BSSV × BSSV modificado
cenário Poligênico, $h^2 = 0,8$

| Informações dos SNPS com efeito médio dos marcadores obtidos pelo método BSSV | | | | Informações dos SNPS com efeito médio dos marcadores obtidos pelo método BSSV modificado | | | |
|---|------------|-----------|-----------------------|--|------------|-----------|-----------------------|
| Marca | Cromossomo | Posição | Efeito médio marcador | Marca | Cromossomo | Posição | Efeito médio marcador |
| <u>5826</u> | 6 | 386200335 | 2,40931143 | <u>5826</u> | 6 | 386200335 | 2,70845813 |
| <u>1562</u> | 2 | 363489782 | 1,55045078 | <u>1562</u> | 2 | 363489782 | 2,57431138 |
| 3517 | 4 | 245955842 | 1,18217590 | <u>3480</u> | 4 | 245802107 | 2,19245491 |
| 3480 | 4 | 245802107 | 0,27898980 | <u>2517</u> | 3 | 380243227 | 1,22414549 |
| <u>5862</u> | 6 | 399021555 | 0,10925866 | <u>4549</u> | 5 | 394247750 | 1,16665011 |
| <u>5560</u> | 6 | 291465765 | 0,08287038 | <u>4629</u> | 5 | 468337830 | 1,11714754 |
| <u>4629</u> | 5 | 468337830 | 0,06794296 | <u>1093</u> | 2 | 20954874 | 0,95155164 |
| <u>2517</u> | 3 | 380243227 | 0,06349708 | 3517 | 4 | 245955842 | 0,81549125 |
| 3736 | 4 | 246865787 | 0,05874724 | <u>3261</u> | 4 | 212893509 | 0,69996686 |
| 4341 | 5 | 210890745 | 0,05435891 | <u>3960</u> | 4 | 247796507 | 0,52115584 |
| 1691 | 2 | 455671386 | 0,04932254 | <u>3288</u> | 4 | 235779015 | 0,48282388 |
| 2736 | 3 | 583965568 | 0,04760805 | <u>5033</u> | 6 | 75187712 | 0,46830289 |
| 3220 | 4 | 180309373 | 0,04674362 | <u>4951</u> | 5 | 766550402 | 0,32934889 |
| 3389 | 4 | 245424002 | 0,03924075 | 3318 | 4 | 245131700 | 0,26783409 |
| 3835 | 4 | 247277132 | 0,03529515 | <u>5862</u> | 6 | 399021555 | 0,22314115 |
| 805 | 1 | 521117995 | 0,03495469 | <u>5560</u> | 6 | 291465765 | 0,22196764 |
| 1865 | 2 | 580009364 | 0,03446140 | 3476 | 4 | 245785487 | 0,17975700 |
| 5636 | 6 | 318532785 | 0,03417793 | 3488 | 4 | 245835347 | 0,15828478 |
| 3760 | 4 | 246965507 | 0,02954686 | <u>5590</u> | 6 | 302150115 | 0,14209589 |
| 5882 | 6 | 406144455 | 0,02936819 | 89 | 1 | 28084996 | 0,13675921 |
| 966 | 1 | 629005068 | 0,02877715 | <u>382</u> | 1 | 262715518 | 0,12773839 |
| 89 | 1 | 28084996 | 0,02589127 | 4085 | 5 | 20200987 | 0,12000420 |
| 4087 | 5 | 21770965 | 0,02537412 | <u>4648</u> | 5 | 485934224 | 0,11033948 |
| 3643 | 4 | 246479372 | 0,02516664 | 5472 | 6 | 260125005 | 0,10079900 |
| 2968 | 3 | 799781016 | 0,02485753 | <u>3337</u> | 4 | 245207942 | 0,09870571 |
| 3550 | 4 | 246092957 | 0,02484530 | <u>5643</u> | 6 | 321025800 | 0,09630312 |
| 5931 | 6 | 423595560 | 0,02481491 | 5109 | 6 | 113385418 | 0,09169950 |
| 3998 | 4 | 247954397 | 0,02436725 | 3406 | 4 | 245494637 | 0,08001256 |
| 4337 | 5 | 209498142 | 0,02433552 | 1528 | 2 | 339193855 | 0,07815269 |
| 5244 | 6 | 178923945 | 0,02427051 | 1686 | 2 | 452098456 | 0,07060968 |
| 3346 | 4 | 245245337 | 0,02379348 | 5461 | 6 | 256207410 | 0,07027736 |
| <u>1093</u> | 2 | 20954874 | 0,02340516 | <u>4849</u> | 5 | 672085550 | 0,06915416 |
| 3535 | 4 | 246030632 | 0,02332603 | 3416 | 4 | 245536187 | 0,06812315 |
| 2840 | 3 | 680710424 | 0,02277077 | 5416 | 6 | 240180885 | 0,06525916 |
| 4453 | 5 | 305339654 | 0,02247994 | 2289 | 3 | 200142461 | 0,06043246 |
| 3301 | 4 | 242019500 | 0,02199768 | 3313 | 4 | 244992633 | 0,05947714 |
| 3632 | 4 | 246433667 | 0,02160994 | <u>2279</u> | 3 | 194643731 | 0,05687205 |
| 2584 | 3 | 442569240 | 0,02148576 | 3032 | 4 | 28192994 | 0,05637520 |
| 4138 | 5 | 62622237 | 0,02077362 | 5417 | 6 | 240537030 | 0,05620728 |
| 4895 | 5 | 714687346 | 0,02021107 | 5131 | 6 | 129927781 | 0,05543521 |
| 5448 | 6 | 251577525 | 0,01999268 | 4593 | 5 | 434997294 | 0,05427564 |
| 3557 | 4 | 246122042 | 0,01989489 | 2901 | 3 | 737455003 | 0,05225533 |
| 3658 | 4 | 246541697 | 0,01972700 | 1249 | 2 | 161879250 | 0,05216266 |
| 5673 | 6 | 331710150 | 0,01956885 | <u>3448</u> | 4 | 245669147 | 0,05089299 |
| <u>1646</u> | 2 | 423515013 | 0,01868875 | 158 | 1 | 56175806 | 0,04900148 |
| 4838 | 5 | 661898164 | 0,01866101 | 3220 | 4 | 180309373 | 0,04763198 |
| 3682 | 4 | 246641417 | 0,01862016 | 105 | 1 | 35621803 | 0,04600655 |
| <u>3261</u> | 4 | 212893509 | 0,01851502 | 3632 | 4 | 246433667 | 0,04574681 |
| <u>4648</u> | 5 | 485934224 | 0,01751170 | 3576 | 4 | 246200987 | 0,04370231 |
| 3773 | 4 | 247019522 | 0,01741209 | 4940 | 5 | 756363016 | 0,04283329 |
| 105 | 1 | 35621803 | 0,01725526 | <u>1646</u> | 2 | 423515013 | 0,04238436 |
| <u>5590</u> | 6 | 302150115 | 0,01716910 | 251 | 1 | 183647544 | 0,04187620 |
| 3919 | 4 | 247626152 | 0,01712842 | 3679 | 4 | 246628952 | 0,04123238 |
| 5990 | 6 | 444608115 | 0,01692844 | 5636 | 6 | 318532785 | 0,04109467 |
| 5653 | 6 | 324587250 | 0,01642825 | 1087 | 2 | 20042233 | 0,03926917 |
| 1686 | 2 | 452098456 | 0,01629602 | 3755 | 4 | 246944732 | 0,03891556 |
| 3045 | 4 | 35769506 | 0,01587553 | 5357 | 6 | 219168330 | 0,03861942 |
| 5365 | 6 | 222017490 | 0,01579512 | 3663 | 4 | 246562472 | 0,03809449 |
| 4117 | 5 | 57831576 | 0,01574040 | 3160 | 4 | 148408488 | 0,03802858 |
| 4649 | 5 | 486860350 | 0,01565660 | 5771 | 6 | 366612360 | 0,03688923 |

marca grifada - associação com gene simulado

Fonte: Da autora (2019)

Tabelas RR-BLUP modificado x GWAS

Tabela 26 – RR-BLUP modificado × GWAS
cenário Oligogênico, $h^2 = 0,2$

| RR-BLUP modificado | | | | | | | | GWAS | | | |
|--------------------|----|-----------|------------|-------|----|-----------|--------------------------------|-------|----|-----------|--------------------------------|
| Marca | Cr | Pos | wald | Marca | Cr | Pos | $-\log_{10}(\text{valor} - p)$ | Marca | Cr | Pos | $-\log_{10}(\text{valor} - p)$ |
| 4004 | 5 | 2123914 | 0,36974271 | 4004 | 5 | 2123914 | 0,26508458 | 4004 | 5 | 2123914 | 7,597,231 |
| 1029 | 2 | 4679573 | 0,15100398 | 1029 | 2 | 4679573 | 0,15640737 | 2983 | 3 | 813734601 | 4,887,828 |
| 2983 | 3 | 813734601 | 0,13074693 | 2983 | 3 | 813734601 | 0,14408196 | 4260 | 5 | 181672656 | 4,265,723 |
| 5521 | 6 | 277576110 | 0,12318400 | 5521 | 6 | 277576110 | 0,13929882 | 1029 | 2 | 4679573 | 3,633,500 |
| 2897 | 3 | 733734047 | 0,12149655 | 2897 | 3 | 733734047 | 0,13821656 | 2896 | 3 | 732803808 | 3,457,414 |
| 4260 | 5 | 181672656 | 0,10887310 | 4260 | 5 | 181672656 | 0,12992940 | 4303 | 5 | 200813183 | 3,316,936 |
| 4621 | 5 | 460928822 | 0,10589688 | 4621 | 5 | 460928822 | 0,12792259 | 4877 | 5 | 698017078 | 3,296,434 |
| 353 | 1 | 240574247 | 0,10578810 | 353 | 1 | 240574247 | 0,12784883 | 2897 | 3 | 733734047 | 3,229,439 |
| 2909 | 3 | 744896915 | 0,10513342 | 2909 | 3 | 744896915 | 0,12740430 | 4047 | 5 | 9965534 | 3,221,998 |
| 5028 | 6 | 57774238 | 0,09911753 | 5028 | 6 | 57774238 | 0,12326777 | 4621 | 5 | 460928822 | 3,214,396 |

Fonte: Da autora (2019)

Tabela 27 – RR-BLUP modificado × GWAS
cenário Oligogênico, $h^2 = 0,5$

| RR-BLUP modificado | | | | | | | | GWAS | | | |
|--------------------|----|-----------|------------|-------|----|-----------|--------------------------------|-------|----|-----------|--------------------------------|
| Marca | Cr | Pos | wald | Marca | Cr | Pos | $-\log_{10}(\text{valor} - p)$ | Marca | Cr | Pos | $-\log_{10}(\text{valor} - p)$ |
| 4004 | 5 | 2123914 | 0,72883655 | 4004 | 5 | 2123914 | 0,4053196 | 4004 | 5 | 2123914 | 13,339721 |
| 3677 | 4 | 246620642 | 0,52639250 | 3677 | 4 | 246620642 | 0,3296359 | 3677 | 4 | 246620642 | 9,102989 |
| 4877 | 5 | 698017078 | 0,43552431 | 4877 | 5 | 698017078 | 0,2930343 | 4877 | 5 | 698017078 | 7,418931 |
| 3805 | 4 | 247152482 | 0,32130487 | 3805 | 4 | 247152482 | 0,2434974 | 3805 | 4 | 247152482 | 5,668896 |
| 5265 | 6 | 186402990 | 0,31103897 | 5265 | 6 | 186402990 | 0,2387916 | 4260 | 5 | 181672656 | 5,507352 |
| 3933 | 4 | 247684322 | 0,28536331 | 3933 | 4 | 247684322 | 0,2267935 | 4625 | 5 | 464633326 | 5,215285 |
| 4260 | 5 | 181672656 | 0,23745980 | 4260 | 5 | 181672656 | 0,2033932 | 3933 | 4 | 247684322 | 4,264209 |
| 3792 | 4 | 247098467 | 0,23696099 | 3792 | 4 | 247098467 | 0,2031415 | 3549 | 4 | 246088802 | 3,997775 |
| 4388 | 5 | 245141464 | 0,23261374 | 4388 | 5 | 245141464 | 0,2009404 | 4772 | 5 | 600773848 | 3,850929 |
| 4625 | 5 | 464633326 | 0,22842801 | 4625 | 5 | 464633326 | 0,1988077 | 3037 | 4 | 29209539 | 3,656472 |

Fonte: Da autora (2019)

Tabela 28 – RR-BLUP modificado × GWAS
cenário Oligogênico, $h^2 = 0,8$

| RR-BLUP modificado | | | | | | | | GWAS | | | |
|--------------------|----|-----------|------------|---|----|-----------|--------------------------------|--------------------------------|----|-----------|--------------------------------|
| wald | | | | $-\log_{10}(\text{valor} - p)$ teste Wald | | | | $-\log_{10}(\text{valor} - p)$ | | | |
| Marca | Cr | Pos | wald | Marca | Cr | Pos | $-\log_{10}(\text{valor} - p)$ | Marca | Cr | Pos | $-\log_{10}(\text{valor} - p)$ |
| 4004 | 5 | 2123914 | 230,227971 | 4004 | 5 | 2123914 | 0,8887904 | 4004 | 5 | 2123914 | 38,278105 |
| 4877 | 5 | 698017078 | 0,73301435 | 4877 | 5 | 698017078 | 0,4068160 | 4877 | 5 | 698017078 | 15,782135 |
| 4303 | 5 | 200813183 | 0,57196102 | 4303 | 5 | 200813183 | 0,3472889 | 4260 | 5 | 181672656 | 10,589633 |
| 4260 | 5 | 181672656 | 0,56244318 | 4260 | 5 | 181672656 | 0,3436358 | 3677 | 4 | 246620642 | 6,920932 |
| 3677 | 4 | 246620642 | 0,52890270 | 3677 | 4 | 246620642 | 0,3306194 | 4303 | 5 | 200813183 | 6,386065 |
| 4772 | 5 | 600773848 | 0,41021571 | 4772 | 5 | 600773848 | 0,2824460 | 4237 | 5 | 171688991 | 5,372977 |
| 5265 | 6 | 186402990 | 0,40090654 | 5265 | 6 | 186402990 | 0,2785015 | 5412 | 6 | 238756305 | 4,982878 |
| 4388 | 5 | 245141464 | 0,37769081 | 4388 | 5 | 245141464 | 0,2685388 | 4625 | 5 | 464633326 | 4,885838 |
| 4621 | 5 | 460928822 | 0,36620210 | 4621 | 5 | 460928822 | 0,2635383 | 4047 | 5 | 9965534 | 4,743689 |
| 4047 | 5 | 9965534 | 0,36463176 | 4047 | 5 | 9965534 | 0,2628511 | 4772 | 5 | 600773848 | 3,910513 |

Fonte: Da autora (2019)

Tabela 29 – RR-BLUP modificado × GWAS
cenário Poligênico, $h^2 = 0,2$

| RR-BLUP modificado | | | | | | | GWAS | | | | |
|--------------------|----|---|------------|-------|--------------------------------|-----------|--------------------------------|-------|----|-----------|--------------------------------|
| wald | | $-\log_{10}(\text{valor} - p)$ teste Wald | | | $-\log_{10}(\text{valor} - p)$ | | | | | | |
| Marca | Cr | Pos | wald | Marca | Cr | Pos | $-\log_{10}(\text{valor} - p)$ | Marca | Cr | Pos | $-\log_{10}(\text{valor} - p)$ |
| 5862 | 6 | 399021555 | 0,19373208 | 5862 | 6 | 399021555 | 0,18056898 | 5826 | 6 | 386200335 | 4,358662 |
| 5826 | 6 | 386200335 | 0,15902119 | 5826 | 6 | 386200335 | 0,16111360 | 3480 | 4 | 245802107 | 4,130539 |
| 5990 | 6 | 444608115 | 0,14044423 | 5990 | 6 | 444608115 | 0,15006503 | 3941 | 4 | 247717562 | 3,956526 |
| 3480 | 4 | 245802107 | 0,12750342 | 3480 | 4 | 245802107 | 0,14204380 | 5862 | 6 | 399021555 | 3,582077 |
| 3941 | 4 | 247717562 | 0,12650063 | 3941 | 4 | 247717562 | 0,14140971 | 4966 | 5 | 780442292 | 3,359942 |
| 5269 | 6 | 187827570 | 0,12016256 | 5269 | 6 | 187827570 | 0,13735693 | 5801 | 6 | 377296710 | 3,308852 |
| 3766 | 4 | 246990437 | 0,11747831 | 3766 | 4 | 246990437 | 0,13561607 | 2840 | 3 | 680710424 | 3,179535 |
| 5239 | 6 | 177143220 | 0,11319816 | 5239 | 6 | 177143220 | 0,13280835 | 5239 | 6 | 177143220 | 3,076366 |
| 4966 | 5 | 780442292 | 0,10819365 | 4966 | 5 | 780442292 | 0,12947315 | 3517 | 4 | 245955842 | 3,035031 |
| 3658 | 4 | 246541697 | 0,10650456 | 3658 | 4 | 246541697 | 0,12833409 | 1408 | 2 | 253443525 | 2,992735 |
| 5801 | 6 | 377296710 | 0,10401444 | 5801 | 6 | 377296710 | 0,12664200 | 5123 | 6 | 121839487 | 2,850738 |
| 3718 | 4 | 246790997 | 0,10195205 | 3718 | 4 | 246790997 | 0,12522862 | 3658 | 4 | 246541697 | 2,838595 |
| 4070 | 5 | 13981667 | 0,09631089 | 4070 | 5 | 13981667 | 0,12130449 | 3810 | 4 | 247173257 | 2,817565 |
| 5728 | 6 | 351298125 | 0,09604768 | 5728 | 6 | 351298125 | 0,12111924 | 805 | 1 | 521117995 | 2,805084 |
| 3126 | 4 | 122795391 | 0,09372637 | 3126 | 4 | 122795391 | 0,11947666 | 3515 | 4 | 245947532 | 2,764010 |
| 3517 | 4 | 245955842 | 0,09077683 | 3517 | 4 | 245955842 | 0,11736626 | 3766 | 4 | 246990437 | 2,738281 |
| 1573 | 2 | 371350229 | 0,09077573 | 1573 | 2 | 371350229 | 0,11736546 | 3420 | 4 | 245552807 | 2,687064 |
| 4960 | 5 | 774885536 | 0,09046394 | 4960 | 5 | 774885536 | 0,11714080 | 5728 | 6 | 351298125 | 2,683066 |
| 5478 | 6 | 262261875 | 0,09002404 | 5478 | 6 | 262261875 | 0,11682332 | 5269 | 6 | 187827570 | 2,602368 |
| 5763 | 6 | 363763200 | 0,08775950 | 5763 | 6 | 363763200 | 0,11517913 | 1562 | 2 | 363489782 | 2,588188 |
| 4582 | 5 | 424809908 | 0,08649482 | 4582 | 5 | 424809908 | 0,11425359 | 2578 | 3 | 436987806 | 2,492446 |
| 3044 | 4 | 35384247 | 0,08311192 | 3044 | 4 | 35384247 | 0,11175101 | 1573 | 2 | 371350229 | 2,436171 |
| 5350 | 6 | 216675315 | 0,08289197 | 5350 | 6 | 216675315 | 0,11158690 | 3718 | 4 | 246790997 | 2,397791 |
| 1408 | 2 | 253443525 | 0,08217105 | 1408 | 2 | 253443525 | 0,11104781 | 3960 | 4 | 247796507 | 2,396755 |
| 842 | 1 | 545911919 | 0,08105469 | 842 | 1 | 545911919 | 0,11020926 | 2025 | 3 | 9124190 | 2,390213 |
| 37 | 1 | 9300541 | 0,08095833 | 37 | 1 | 9300541 | 0,11013666 | 1941 | 2 | 634317906 | 2,383527 |
| 989 | 1 | 644417507 | 0,08050032 | 989 | 1 | 644417507 | 0,10979113 | 4582 | 5 | 424809908 | 2,381976 |
| 2616 | 3 | 472336888 | 0,07918385 | 2616 | 3 | 472336888 | 0,10879355 | 1268 | 2 | 169989408 | 2,359644 |
| 4838 | 5 | 661898164 | 0,07792919 | 4838 | 5 | 661898164 | 0,10783660 | 5763 | 6 | 363763200 | 2,356424 |
| 5673 | 6 | 331710150 | 0,07780183 | 5673 | 6 | 331710150 | 0,10773911 | 5933 | 6 | 424307850 | 2,329228 |
| 699 | 1 | 450086755 | 0,07714023 | 699 | 1 | 450086755 | 0,10723167 | 4070 | 5 | 13981667 | 2,307127 |
| 4862 | 5 | 684125188 | 0,07672836 | 4862 | 5 | 684125188 | 0,10691489 | 5990 | 6 | 444608115 | 2,297095 |
| 3964 | 4 | 247813127 | 0,07412537 | 3964 | 4 | 247813127 | 0,10489677 | 842 | 1 | 545911919 | 2,284275 |
| 629 | 1 | 403179332 | 0,07396210 | 629 | 1 | 403179332 | 0,110476924 | 5954 | 6 | 431786895 | 2,269155 |
| 1268 | 2 | 169989408 | 0,07269171 | 1268 | 2 | 169989408 | 0,10377303 | 382 | 1 | 262715518 | 2,253707 |
| 933 | 1 | 606891568 | 0,07216959 | 933 | 1 | 606891568 | 0,10336155 | 5478 | 6 | 262261875 | 2,213185 |
| 1935 | 2 | 630030390 | 0,07210487 | 1935 | 2 | 630030390 | 0,10331046 | 5350 | 6 | 216675315 | 2,187097 |
| 3354 | 4 | 245278577 | 0,07178807 | 3354 | 4 | 245278577 | 0,10306012 | 2726 | 3 | 574663178 | 2,168420 |
| 1600 | 2 | 390644053 | 0,07139405 | 1600 | 2 | 390644053 | 0,10274813 | 5693 | 6 | 338833050 | 2,146788 |
| 2541 | 3 | 402568963 | 0,07063480 | 2541 | 3 | 402568963 | 0,10214497 | 3296 | 4 | 239237605 | 2,138550 |
| 3881 | 4 | 247468262 | 0,07060333 | 3881 | 4 | 247468262 | 0,10211991 | 4960 | 5 | 774885536 | 2,124039 |
| 5818 | 6 | 383351175 | 0,07039362 | 5818 | 6 | 383351175 | 0,10195282 | 3399 | 4 | 245465552 | 2,105284 |
| 5789 | 6 | 373022970 | 0,07036639 | 5789 | 6 | 373022970 | 0,10193111 | 2901 | 3 | 737455003 | 2,105209 |
| 1617 | 2 | 402792016 | 0,06954492 | 1617 | 2 | 402792016 | 0,10127453 | 4365 | 5 | 223840566 | 2,102395 |
| 4403 | 5 | 259033354 | 0,06916725 | 4403 | 5 | 259033354 | 0,10097161 | 4087 | 5 | 21770965 | 2,099218 |
| 2578 | 3 | 436987806 | 0,06905980 | 2578 | 3 | 436987806 | 0,10088530 | 933 | 1 | 606891568 | 2,090111 |
| 5379 | 6 | 227003520 | 0,06780605 | 5379 | 6 | 227003520 | 0,09987419 | 3044 | 4 | 35384247 | 2,088211 |
| 3164 | 4 | 149588842 | 0,06743002 | 3164 | 4 | 149588842 | 0,09956946 | 620 | 1 | 397148378 | 2,085016 |
| 4961 | 5 | 775811662 | 0,06692482 | 4961 | 5 | 775811662 | 0,09915895 | 4838 | 5 | 661898164 | 2,083066 |
| 3396 | 4 | 245453087 | 0,06673877 | 3396 | 4 | 245453087 | 0,09900745 | 5175 | 6 | 151958886 | 2,081309 |
| 3885 | 4 | 247484882 | 0,06628163 | 3885 | 4 | 247484882 | 0,09863448 | 2480 | 3 | 345824384 | 2,068159 |
| 3454 | 4 | 245694077 | 0,06620288 | 3454 | 4 | 245694077 | 0,09857012 | 3532 | 4 | 246018167 | 2,065721 |
| 825 | 1 | 534520116 | 0,06589351 | 825 | 1 | 534520116 | 0,09831699 | 5789 | 6 | 373022970 | 2,064492 |
| 5277 | 6 | 190676730 | 0,06584033 | 5277 | 6 | 190676730 | 0,09827343 | 2584 | 3 | 442569240 | 2,058015 |
| 1562 | 2 | 363489782 | 0,06581093 | 1562 | 2 | 363489782 | 0,09824934 | 3881 | 4 | 247468262 | 2,057619 |
| 3515 | 4 | 245947532 | 0,06465032 | 3515 | 4 | 245947532 | 0,09729486 | 5756 | 6 | 361270185 | 2,037824 |
| 3377 | 4 | 245374142 | 0,06464013 | 3377 | 4 | 245374142 | 0,09728645 | 1657 | 2 | 431375459 | 2,033325 |
| 65 | 1 | 17601517 | 0,06441923 | 65 | 1 | 17601517 | 0,09710398 | 4604 | 5 | 445184680 | 2,018235 |
| 4023 | 5 | 6820571 | 0,06374131 | 4023 | 5 | 6820571 | 0,09654238 | 2968 | 3 | 799781016 | 2,013910 |

Fonte: Da autora (2019)

Tabela 30 – RR-BLUP modificado × GWAS
cenário Poligênico, $h^2 = 0,5$

| RR-BLUP modificado | | | | | | | | GWAS | | | |
|--------------------|----|-----------|------------|---|----|-----------|--------------------------------|--------------------------------|----|-----------|--------------------------------|
| wald | | | | $-\log_{10}(\text{valor} - p)$ teste Wald | | | | $-\log_{10}(\text{valor} - p)$ | | | |
| Marca | Cr | Pos | wald | Marca | Cr | Pos | $-\log_{10}(\text{valor} - p)$ | Marca | Cr | Pos | $-\log_{10}(\text{valor} - p)$ |
| 5826 | 6 | 386200335 | 0,46399464 | 5826 | 6 | 386200335 | 0,3047259 | 5826 | 6 | 386200335 | 8,459814 |
| 3480 | 4 | 245802107 | 0,31526315 | 3480 | 4 | 245802107 | 0,2407339 | 3480 | 4 | 245802107 | 6,266500 |
| 5673 | 6 | 331710150 | 0,30333446 | 5673 | 6 | 331710150 | 0,2352266 | 1562 | 2 | 363489782 | 5,390297 |
| 3517 | 4 | 245955842 | 0,27585947 | 3517 | 4 | 245955842 | 0,2222630 | 3261 | 4 | 212893509 | 4,081755 |
| 5277 | 6 | 190676730 | 0,25953142 | 5277 | 6 | 190676730 | 0,2143554 | 5673 | 6 | 331710150 | 3,983664 |
| 5862 | 6 | 399021555 | 0,25568590 | 5862 | 6 | 399021555 | 0,2124690 | 3517 | 4 | 245955842 | 3,790299 |
| 5846 | 6 | 393323235 | 0,23030974 | 5846 | 6 | 393323235 | 0,1997682 | 5846 | 6 | 393323235 | 3,772569 |
| 5884 | 6 | 406856745 | 0,21135398 | 5884 | 6 | 406856745 | 0,1899637 | 2517 | 3 | 380243227 | 3,400196 |
| 4337 | 5 | 209498142 | 0,20910021 | 4337 | 5 | 209498142 | 0,1887780 | 5933 | 6 | 424307850 | 3,338913 |
| 3261 | 4 | 212893509 | 0,20363803 | 3261 | 4 | 212893509 | 0,1858855 | 5277 | 6 | 190676730 | 3,331761 |
| 1562 | 2 | 363489782 | 0,20216683 | 1562 | 2 | 363489782 | 0,1851018 | 1646 | 2 | 423515013 | 3,277031 |
| 4838 | 5 | 661898164 | 0,18870665 | 4838 | 5 | 661898164 | 0,1778350 | 3736 | 4 | 246865787 | 3,273612 |
| 4317 | 5 | 204516414 | 0,18527275 | 4317 | 5 | 204516414 | 0,1759520 | 5884 | 6 | 406856745 | 3,248133 |
| 5763 | 6 | 363763200 | 0,17708167 | 5763 | 6 | 363763200 | 0,1714089 | 4317 | 5 | 204516414 | 3,192595 |
| 1093 | 2 | 20954874 | 0,16656456 | 1093 | 2 | 20954874 | 0,1654628 | 4649 | 5 | 486860350 | 3,177421 |
| 4290 | 5 | 193521948 | 0,16410165 | 4290 | 5 | 193521948 | 0,1640509 | 2749 | 3 | 596058675 | 3,177365 |
| 4649 | 5 | 486860350 | 0,16377753 | 4649 | 5 | 486860350 | 0,1638645 | 5862 | 6 | 399021555 | 3,117959 |
| 5619 | 6 | 312478320 | 0,16043453 | 5619 | 6 | 312478320 | 0,1619341 | 3381 | 4 | 245390762 | 3,035844 |
| 1865 | 2 | 580009364 | 0,15709997 | 1865 | 2 | 580009364 | 0,1599939 | 5670 | 6 | 330641715 | 3,008389 |
| 1646 | 2 | 423515013 | 0,15497289 | 1646 | 2 | 423515013 | 0,1587484 | 187 | 1 | 83595286 | 2,927185 |
| 4977 | 5 | 790629678 | 0,15277369 | 4977 | 5 | 790629678 | 0,1574540 | 5799 | 6 | 376584420 | 2,906733 |
| 4913 | 5 | 731357614 | 0,14472988 | 4913 | 5 | 731357614 | 0,1526597 | 1093 | 2 | 20954874 | 2,879809 |
| 2749 | 3 | 596058675 | 0,14384773 | 2749 | 3 | 596058675 | 0,1521280 | 1865 | 2 | 580009364 | 2,858210 |
| 2072 | 3 | 53549251 | 0,14188236 | 2072 | 3 | 53549251 | 0,1509390 | 3389 | 4 | 245424002 | 2,708215 |
| 4265 | 5 | 184588958 | 0,14179855 | 4265 | 5 | 184588958 | 0,1508881 | 5033 | 6 | 75187712 | 2,700129 |
| 4966 | 5 | 780442292 | 0,14144356 | 4966 | 5 | 780442292 | 0,1506727 | 5805 | 6 | 378721290 | 2,686021 |
| 3133 | 4 | 128632208 | 0,13935559 | 3133 | 4 | 128632208 | 0,1494013 | 427 | 1 | 268373293 | 2,590842 |
| 5442 | 6 | 249440655 | 0,13871755 | 5442 | 6 | 249440655 | 0,1490113 | 4290 | 5 | 193521948 | 2,568202 |
| 5547 | 6 | 286835880 | 0,13866663 | 5547 | 6 | 286835880 | 0,1489802 | 945 | 1 | 614932841 | 2,566545 |
| 2602 | 3 | 459313542 | 0,13787173 | 2602 | 3 | 459313542 | 0,1484934 | 5698 | 6 | 340613775 | 2,563206 |
| 5244 | 6 | 178923945 | 0,13737406 | 5244 | 6 | 178923945 | 0,1481880 | 4265 | 5 | 184588958 | 2,491110 |
| 5853 | 6 | 395816250 | 0,13734738 | 5853 | 6 | 395816250 | 0,1481717 | 4838 | 5 | 661898164 | 2,438240 |
| 3992 | 4 | 247929467 | 0,13674050 | 3992 | 4 | 247929467 | 0,1477988 | 4453 | 5 | 305339654 | 2,422658 |
| 5698 | 6 | 340613775 | 0,13585008 | 5698 | 6 | 340613775 | 0,1472505 | 1095 | 2 | 20958307 | 2,421069 |
| 4453 | 5 | 305339654 | 0,13252426 | 4453 | 5 | 305339654 | 0,1451907 | 3133 | 4 | 128632208 | 2,401665 |
| 4126 | 5 | 59300236 | 0,13212095 | 4126 | 5 | 59300236 | 0,1449396 | 3096 | 4 | 76123039 | 2,394804 |
| 945 | 1 | 614932841 | 0,13047766 | 945 | 1 | 614932841 | 0,1439135 | 5619 | 6 | 312478320 | 2,369837 |
| 3009 | 4 | 2639015 | 0,13046758 | 3009 | 4 | 2639015 | 0,1439072 | 2877 | 3 | 715129267 | 2,355939 |
| 2517 | 3 | 380243227 | 0,12985646 | 2517 | 3 | 380243227 | 0,1435243 | 3971 | 4 | 247842212 | 2,332747 |
| 4081 | 5 | 19973330 | 0,12796740 | 4081 | 5 | 19973330 | 0,1423365 | 4802 | 5 | 628557628 | 2,332568 |
| 5033 | 6 | 75187712 | 0,12414793 | 5033 | 6 | 75187712 | 0,1399145 | 4648 | 5 | 485934224 | 2,326017 |
| 5361 | 6 | 220592910 | 0,12332797 | 5361 | 6 | 220592910 | 0,1393909 | 5244 | 6 | 178923945 | 2,324638 |
| 3633 | 4 | 246437822 | 0,12257953 | 3633 | 4 | 246437822 | 0,1389118 | 1358 | 2 | 214423504 | 2,304391 |
| 4170 | 5 | 75234294 | 0,12239200 | 4170 | 5 | 75234294 | 0,1387916 | 3972 | 4 | 247846367 | 2,286949 |
| 499 | 1 | 291031959 | 0,12228330 | 499 | 1 | 291031959 | 0,1387219 | 499 | 1 | 291031959 | 2,284529 |
| 3224 | 4 | 184373943 | 0,12067410 | 3224 | 4 | 184373943 | 0,1376870 | 1185 | 2 | 58354695 | 2,283718 |
| 5799 | 6 | 376584420 | 0,12011989 | 5799 | 6 | 376584420 | 0,1373294 | 5103 | 6 | 113384656 | 2,265262 |
| 1210 | 2 | 128180134 | 0,11937336 | 1210 | 2 | 128180134 | 0,1368466 | 1321 | 2 | 198757428 | 2,264077 |
| 1417 | 2 | 259874800 | 0,11903382 | 1417 | 2 | 259874800 | 0,1366267 | 821 | 1 | 531839692 | 2,263088 |
| 3864 | 4 | 247397627 | 0,11786230 | 3864 | 4 | 247397627 | 0,1358660 | 308 | 1 | 216989434 | 2,252107 |
| 5933 | 6 | 424307850 | 0,11451911 | 5933 | 6 | 424307850 | 0,1336792 | 1354 | 2 | 214263400 | 2,231001 |
| 427 | 1 | 268373293 | 0,11382881 | 427 | 1 | 268373293 | 0,1332246 | 5547 | 6 | 286835880 | 2,222143 |
| 1354 | 2 | 214263400 | 0,11355496 | 1354 | 2 | 214263400 | 0,1330439 | 4629 | 5 | 468337830 | 2,186507 |
| 4554 | 5 | 398878380 | 0,11345289 | 4554 | 5 | 398878380 | 0,1329766 | 4170 | 5 | 75234294 | 2,178396 |
| 1358 | 2 | 214423504 | 0,11173461 | 1358 | 2 | 214423504 | 0,1318390 | 3992 | 4 | 247929467 | 2,174125 |
| 1263 | 2 | 169265278 | 0,11146475 | 1263 | 2 | 169265278 | 0,1316597 | 3005 | 4 | 2637306 | 2,168391 |
| 1185 | 2 | 58354695 | 0,11125751 | 1185 | 2 | 58354695 | 0,1315219 | 2072 | 3 | 53549251 | 2,162393 |
| 3960 | 4 | 247796507 | 0,11017523 | 3960 | 4 | 247796507 | 0,1308007 | 1417 | 2 | 259874800 | 2,154835 |
| 1631 | 2 | 412796221 | 0,10960652 | 1631 | 2 | 412796221 | 0,1304207 | 5789 | 6 | 373022970 | 2,123864 |

Fonte: Da autora (2019)

Tabela 31 – RR-BLUP modificado × GWAS
cenário Poligênico, $h^2 = 0,8$

| RR-BLUP modificado | | | | | | | GWAS | | | | |
|--------------------|----|-----------|---|-------|----|-----------|--------------------------------|-------|----|-----------|--------------------------------|
| wald | | | $-\log_{10}(\text{valor} - p)$ teste Wald | | | | $-\log_{10}(\text{valor} - p)$ | | | | |
| Marca | Cr | Pos | wald | Marca | Cr | Pos | $-\log_{10}(\text{valor} - p)$ | Marca | Cr | Pos | $-\log_{10}(\text{valor} - p)$ |
| 5826 | 6 | 386200335 | 0,83392365 | 5826 | 6 | 386200335 | 0,4423239 | 5826 | 6 | 386200335 | 11,587916 |
| 3480 | 4 | 245802107 | 0,38542557 | 3480 | 4 | 245802107 | 0,2718786 | 3480 | 4 | 245802107 | 8,318912 |
| 1562 | 2 | 363489782 | 0,35556658 | 1562 | 2 | 363489782 | 0,2588651 | 5954 | 6 | 431786895 | 5,722483 |
| 5862 | 6 | 399021555 | 0,32011287 | 5862 | 6 | 399021555 | 0,2429535 | 1562 | 2 | 363489782 | 5,662491 |
| 3517 | 4 | 245955842 | 0,31901940 | 3517 | 4 | 245955842 | 0,2424540 | 5033 | 6 | 75187712 | 5,350154 |
| 5954 | 6 | 431786895 | 0,29506895 | 5954 | 6 | 431786895 | 0,2313690 | 5862 | 6 | 399021555 | 4,623745 |
| 5033 | 6 | 75187712 | 0,24298537 | 5033 | 6 | 75187712 | 0,2061690 | 2517 | 3 | 380243227 | 4,065757 |
| 3658 | 4 | 246541697 | 0,24034733 | 3658 | 4 | 246541697 | 0,2048465 | 5590 | 6 | 302150115 | 3,514904 |
| 5763 | 6 | 363763200 | 0,23812002 | 5763 | 6 | 363763200 | 0,2037260 | 3736 | 4 | 246865787 | 3,498160 |
| 5801 | 6 | 377296710 | 0,22878809 | 5801 | 6 | 377296710 | 0,1989917 | 1093 | 2 | 20954874 | 3,365941 |
| 3998 | 4 | 247954397 | 0,21928275 | 3998 | 4 | 247954397 | 0,1941004 | 2292 | 3 | 200201258 | 3,072857 |
| 5846 | 6 | 393323235 | 0,21789303 | 5846 | 6 | 393323235 | 0,1933791 | 3517 | 4 | 245955842 | 3,015800 |
| 4649 | 5 | 486860350 | 0,21775947 | 4649 | 5 | 486860350 | 0,1933097 | 859 | 1 | 557303721 | 2,967605 |
| 3261 | 4 | 212893509 | 0,21279019 | 3261 | 4 | 212893509 | 0,1907170 | 3643 | 4 | 246479372 | 2,908547 |
| 1646 | 2 | 423515013 | 0,21248628 | 1646 | 2 | 423515013 | 0,1905577 | 3996 | 4 | 247946087 | 2,879696 |
| 1865 | 2 | 580009364 | 0,21051774 | 1865 | 2 | 580009364 | 0,1895242 | 89 | 1 | 28084996 | 2,879099 |
| 4629 | 5 | 468337830 | 0,20952331 | 4629 | 5 | 468337830 | 0,1890009 | 3918 | 4 | 247621997 | 2,809284 |
| 3810 | 4 | 247173257 | 0,20484550 | 3810 | 4 | 247173257 | 0,1865272 | 5134 | 6 | 131576099 | 2,790126 |
| 2517 | 3 | 380243227 | 0,20436087 | 2517 | 3 | 380243227 | 0,1862698 | 1865 | 2 | 580009364 | 2,786465 |
| 1686 | 2 | 452098456 | 0,19946985 | 1686 | 2 | 452098456 | 0,1836598 | 4629 | 5 | 468337830 | 2,781031 |
| 3996 | 4 | 247946087 | 0,19743498 | 3996 | 4 | 247946087 | 0,1825673 | 5636 | 6 | 318532785 | 2,764711 |
| 966 | 1 | 629005068 | 0,19651010 | 966 | 1 | 629005068 | 0,1820695 | 1304 | 2 | 189293647 | 2,700383 |
| 5442 | 6 | 249440655 | 0,19440838 | 5442 | 6 | 249440655 | 0,1809350 | 5186 | 6 | 157129197 | 2,698615 |
| 3918 | 4 | 247621997 | 0,18797215 | 3918 | 4 | 247621997 | 0,1774333 | 5063 | 6 | 83064740 | 2,698013 |
| 3643 | 4 | 246479372 | 0,18657362 | 3643 | 4 | 246479372 | 0,1766668 | 5805 | 6 | 378721290 | 2,687730 |
| 3760 | 4 | 246965507 | 0,18650130 | 3760 | 4 | 246965507 | 0,1766271 | 3476 | 4 | 245785487 | 2,630147 |
| 382 | 1 | 262715518 | 0,18511485 | 382 | 1 | 262715518 | 0,1758651 | 1646 | 2 | 423515013 | 2,593889 |
| 5698 | 6 | 340613775 | 0,18461930 | 5698 | 6 | 340613775 | 0,1755922 | 5801 | 6 | 377296710 | 2,582850 |
| 5774 | 6 | 367680795 | 0,18376349 | 5774 | 6 | 367680795 | 0,1751204 | 2840 | 3 | 680710424 | 2,569747 |
| 5673 | 6 | 331710150 | 0,18349580 | 5673 | 6 | 331710150 | 0,1749726 | 966 | 1 | 629005068 | 2,547537 |
| 3220 | 4 | 180309373 | 0,18214714 | 3220 | 4 | 180309373 | 0,1742271 | 5530 | 6 | 280781415 | 2,532024 |
| 5171 | 6 | 149251173 | 0,17883314 | 5171 | 6 | 149251173 | 0,1723865 | 4459 | 5 | 310896410 | 2,531509 |
| 2901 | 3 | 737455003 | 0,17816333 | 2901 | 3 | 737455003 | 0,1720131 | 5416 | 6 | 240180885 | 2,488117 |
| 5186 | 6 | 157129197 | 0,17699274 | 5186 | 6 | 157129197 | 0,1713591 | 1933 | 2 | 628601218 | 2,436105 |
| 4265 | 5 | 184588958 | 0,17553741 | 4265 | 5 | 184588958 | 0,1705440 | 3658 | 4 | 246541697 | 2,433635 |
| 3389 | 4 | 245424002 | 0,17108602 | 3389 | 4 | 245424002 | 0,1680353 | 3538 | 4 | 246043097 | 2,428786 |
| 4838 | 5 | 661898164 | 0,16931553 | 4838 | 5 | 661898164 | 0,1670310 | 5442 | 6 | 249440655 | 2,404876 |
| 1178 | 2 | 51785620 | 0,16812515 | 1178 | 2 | 51785620 | 0,1663535 | 3220 | 4 | 180309373 | 2,391637 |
| 5882 | 6 | 406144455 | 0,16770179 | 5882 | 6 | 406144455 | 0,1661122 | 5290 | 6 | 195306615 | 2,384984 |
| 5357 | 6 | 219168330 | 0,16604142 | 5357 | 6 | 219168330 | 0,1651635 | 382 | 1 | 262715518 | 2,379969 |
| 5884 | 6 | 406856745 | 0,16445897 | 5884 | 6 | 406856745 | 0,1642562 | 5782 | 6 | 370529955 | 2,373650 |
| 2840 | 3 | 680710424 | 0,16242791 | 2840 | 3 | 680710424 | 0,1630869 | 1086 | 2 | 19837545 | 2,371760 |
| 5853 | 6 | 395816250 | 0,16228537 | 5853 | 6 | 395816250 | 0,1630046 | 4549 | 5 | 394247750 | 2,352150 |
| 5683 | 6 | 335271600 | 0,15943990 | 5683 | 6 | 335271600 | 0,1613570 | 2998 | 3 | 827688186 | 2,352145 |
| 2998 | 3 | 827688186 | 0,15892112 | 2998 | 3 | 827688186 | 0,1610554 | 1880 | 2 | 590728155 | 2,342969 |
| 5619 | 6 | 312478320 | 0,15855931 | 5619 | 6 | 312478320 | 0,1608449 | 5698 | 6 | 340613775 | 2,338420 |
| 5982 | 6 | 441758955 | 0,15649454 | 5982 | 6 | 441758955 | 0,1596400 | 3295 | 4 | 238553534 | 2,331521 |
| 1079 | 2 | 14751136 | 0,15641094 | 1079 | 2 | 14751136 | 0,1595911 | 5171 | 6 | 149251173 | 2,318388 |
| 859 | 1 | 557303721 | 0,15638225 | 859 | 1 | 557303721 | 0,1595743 | 3337 | 4 | 245207942 | 2,312483 |
| 3301 | 4 | 242019500 | 0,15637905 | 3301 | 4 | 242019500 | 0,1595725 | 728 | 1 | 469519830 | 2,297564 |
| 3515 | 4 | 245947532 | 0,15593786 | 3515 | 4 | 245947532 | 0,1593142 | 1686 | 2 | 452098456 | 2,294555 |
| 5547 | 6 | 286835880 | 0,15535544 | 5547 | 6 | 286835880 | 0,1589728 | 3632 | 4 | 246433667 | 2,263630 |
| 1661 | 2 | 434233804 | 0,15533048 | 1661 | 2 | 434233804 | 0,1589582 | 3261 | 4 | 212893509 | 2,223022 |
| 4476 | 5 | 326640552 | 0,15421566 | 4476 | 5 | 326640552 | 0,1583035 | 5641 | 6 | 320313510 | 2,207166 |
| 1759 | 2 | 504263240 | 0,15318140 | 1759 | 2 | 504263240 | 0,1576945 | 5244 | 6 | 178923945 | 2,205616 |
| 3476 | 4 | 245785487 | 0,15233832 | 3476 | 4 | 245785487 | 0,1571969 | 4849 | 5 | 672085550 | 2,204546 |
| 89 | 1 | 28084996 | 0,15169949 | 89 | 1 | 28084996 | 0,1568192 | 1249 | 2 | 161879250 | 2,186632 |
| 5811 | 6 | 380858160 | 0,14741785 | 5811 | 6 | 380858160 | 0,1542725 | 5336 | 6 | 211689285 | 2,170129 |
| 5134 | 6 | 131576099 | 0,14368560 | 5134 | 6 | 131576099 | 0,1520301 | 4394 | 5 | 250698220 | 2,163688 |

Fonte: Da autora (2019)

Tabelas Bayes A modificado x GWAS

Tabela 32 – Bayes A modificado \times GWAS
cenário Oligogênico, $h^2 = 0,2$

| Bayes A modificado | | | | | | | | GWAS | | | |
|--------------------|----|-----------|-------------|-------|----|-----------|--------------------------------|-------|----|-----------|--------------------------------|
| Marca | Cr | Pos | wald | Marca | Cr | Pos | $-\log_{10}(\text{valor} - p)$ | Marca | Cr | Pos | $-\log_{10}(\text{valor} - p)$ |
| 4004 | 5 | 2123914 | 24,12597397 | 4004 | 5 | 2123914 | 6,0446252 | 4004 | 5 | 2123914 | 7,597231 |
| 2983 | 3 | 813734601 | 3,77198286 | 2983 | 3 | 813734601 | 1,2830138 | 2983 | 3 | 813734601 | 4,887828 |
| 1029 | 2 | 4679573 | 1,27128384 | 1029 | 2 | 4679573 | 0,5858192 | 4260 | 5 | 181672656 | 4,265723 |
| 1826 | 2 | 552140507 | 0,87739936 | 1826 | 2 | 552140507 | 0,4572804 | 1029 | 2 | 4679573 | 3,633500 |
| 2897 | 3 | 733734047 | 0,76303412 | 2897 | 3 | 733734047 | 0,4175041 | 2896 | 3 | 732803808 | 3,457414 |
| 4877 | 5 | 698017078 | 0,73238788 | 4877 | 5 | 698017078 | 0,4065918 | 4303 | 5 | 200813183 | 3,316936 |
| 279 | 1 | 202620610 | 0,67801044 | 279 | 1 | 202620610 | 0,3869274 | 4877 | 5 | 698017078 | 3,296434 |
| 2256 | 3 | 186580149 | 0,60824865 | 2256 | 3 | 186580149 | 0,3610637 | 2897 | 3 | 733734047 | 3,229439 |
| 1051 | 2 | 7482709 | 0,55967054 | 1051 | 2 | 7482709 | 0,3425683 | 4047 | 5 | 9965534 | 3,221998 |
| 922 | 1 | 599520402 | 0,41998412 | 922 | 1 | 599520402 | 0,2865557 | 4621 | 5 | 460928822 | 3,214396 |

Fonte: Da autora (2019)

Tabela 33 – Bayes A modificado \times GWAS
cenário Oligogênico, $h^2 = 0,5$

| Bayes A modificado | | | | | | | | GWAS | | | |
|--------------------|----|-----------|-------------|-------|----|-----------|--------------------------------|-------|----|-----------|--------------------------------|
| Marca | Cr | Pos | wald | Marca | Cr | Pos | $-\log_{10}(\text{valor} - p)$ | Marca | Cr | Pos | $-\log_{10}(\text{valor} - p)$ |
| 4004 | 5 | 2123914 | 58,23529102 | 4004 | 5 | 2123914 | 13,6334475 | 4004 | 5 | 2123914 | 13,339721 |
| 3677 | 4 | 246620642 | 47,01928872 | 3677 | 4 | 246620642 | 11,1530971 | 3677 | 4 | 246620642 | 9,102989 |
| 4877 | 5 | 698017078 | 40,18738277 | 4877 | 5 | 698017078 | 9,6368891 | 4877 | 5 | 698017078 | 7,418931 |
| 193 | 1 | 88380847 | 2,51212274 | 193 | 1 | 88380847 | 0,9470228 | 3805 | 4 | 247152482 | 5,668896 |
| 833 | 1 | 539880964 | 0,61552472 | 833 | 1 | 539880964 | 0,3637978 | 4260 | 5 | 181672656 | 5,507352 |
| 538 | 1 | 298413091 | 0,49085916 | 538 | 1 | 298413091 | 0,3155637 | 4625 | 5 | 464633326 | 5,215285 |
| 877 | 1 | 569365630 | 0,48912707 | 877 | 1 | 569365630 | 0,3148703 | 3933 | 4 | 247684322 | 4,264209 |
| 625 | 1 | 400498908 | 0,37183414 | 625 | 1 | 400498908 | 0,2659957 | 3549 | 4 | 246088802 | 3,997775 |
| 186 | 1 | 82563962 | 0,36931427 | 186 | 1 | 82563962 | 0,2648977 | 4772 | 5 | 600773848 | 3,850929 |
| 139 | 1 | 47280990 | 0,34579381 | 139 | 1 | 47280990 | 0,2545318 | 3037 | 4 | 29209539 | 3,656472 |

Fonte: Da autora (2019)

Tabela 34 – Bayes A modificado \times GWAS
cenário Oligogênico, $h^2 = 0,8$

| Bayes A modificado | | | | | | | | GWAS | | | |
|--------------------|----|-----------|--------------|-------|----|-----------|--------------------------------|-------|----|-----------|--------------------------------|
| Marca | Cr | Pos | wald | Marca | Cr | Pos | $-\log_{10}(\text{valor} - p)$ | Marca | Cr | Pos | $-\log_{10}(\text{valor} - p)$ |
| 4004 | 5 | 2123914 | 219,99960211 | 4004 | 5 | 2123914 | 4,904352962 | 4004 | 5 | 2123914 | 38,278105 |
| 4877 | 5 | 698017078 | 104,61597517 | 4877 | 5 | 698017078 | 2,382898581 | 4877 | 5 | 698017078 | 15,782135 |
| 3677 | 4 | 246620642 | 55,15124780 | 3677 | 4 | 246620642 | 1,295232493 | 4260 | 5 | 181672656 | 10,589633 |
| 538 | 1 | 298413091 | 9,89295989 | 538 | 1 | 298413091 | 2,78012145 | 3677 | 4 | 246620642 | 6,920932 |
| 2088 | 3 | 69892547 | 4,49732010 | 2088 | 3 | 69892547 | 1,46918558 | 4303 | 5 | 200813183 | 6,386065 |
| 2000 | 2 | 676478485 | 1,03162515 | 2000 | 2 | 676478485 | 0,50895045 | 4237 | 5 | 171688991 | 5,372977 |
| 4927 | 5 | 744323378 | 0,50354395 | 4927 | 5 | 744323378 | 0,32062020 | 5412 | 6 | 238756305 | 4,982878 |
| 2803 | 3 | 646291581 | 0,42907812 | 2803 | 3 | 646291581 | 0,29035558 | 4625 | 5 | 464633326 | 4,885838 |
| 3315 | 4 | 245051599 | 0,39480090 | 3315 | 4 | 245051599 | 0,27589905 | 4047 | 5 | 9965534 | 4,743689 |
| 900 | 1 | 584778069 | 0,36164799 | 900 | 1 | 584778069 | 0,26154258 | 4772 | 5 | 600773848 | 3,910513 |

Fonte: Da autora (2019)

Tabela 35 – Bayes A modificado \times GWAS
cenário Poligênico, $h^2 = 0,2$

| Bayes A modificado | | | | GWAS | | | | | | | |
|--------------------|----|-----------|------------|-------|----|-----------|--------------------------------|-------|----|-----------|--------------------------------|
| Marca | Cr | Pos | wald | Marca | Cr | Pos | $-\log_{10}(\text{valor} - p)$ | Marca | Cr | Pos | $-\log_{10}(\text{valor} - p)$ |
| 5826 | 6 | 386200335 | 5,90361203 | 5826 | 6 | 386200335 | 1,8207394 | 5826 | 6 | 386200335 | 4,358662 |
| 3941 | 4 | 247717562 | 3,18197902 | 3941 | 4 | 247717562 | 1,1281090 | 3480 | 4 | 245802107 | 4,130539 |
| 3480 | 4 | 245802107 | 2,70114261 | 3480 | 4 | 245802107 | 0,9988014 | 3941 | 4 | 247717562 | 3,956526 |
| 5862 | 6 | 399021555 | 1,42614053 | 5862 | 6 | 399021555 | 0,6337726 | 5862 | 6 | 399021555 | 3,582077 |
| 3718 | 4 | 246790997 | 0,91098114 | 3718 | 4 | 246790997 | 0,4687073 | 4966 | 5 | 780442292 | 3,359942 |
| 1408 | 2 | 253443525 | 0,88753596 | 1408 | 2 | 253443525 | 0,4607408 | 5801 | 6 | 377296710 | 3,308852 |
| 2840 | 3 | 680710424 | 0,83641039 | 2840 | 3 | 680710424 | 0,4500835 | 2840 | 3 | 680710424 | 3,179535 |
| 5269 | 6 | 187827570 | 0,85618610 | 5269 | 6 | 187827570 | 0,4500064 | 5239 | 6 | 177143220 | 3,076366 |
| 5801 | 6 | 377296710 | 0,57251687 | 5801 | 6 | 377296710 | 0,3475018 | 3517 | 4 | 245955842 | 3,035031 |
| 1941 | 2 | 634317906 | 0,50005952 | 1941 | 2 | 634317906 | 0,3192350 | 1408 | 2 | 253443525 | 2,992735 |
| 805 | 1 | 521117995 | 0,47178076 | 805 | 1 | 521117995 | 0,3078857 | 5123 | 6 | 121839487 | 2,850738 |
| 3420 | 4 | 245552807 | 0,46973535 | 3420 | 4 | 245552807 | 0,3070571 | 3658 | 4 | 246541697 | 2,838595 |
| 3515 | 4 | 245947532 | 0,45569680 | 3515 | 4 | 245947532 | 0,3013411 | 3810 | 4 | 247173257 | 2,817565 |
| 1573 | 2 | 371350229 | 0,43912687 | 1573 | 2 | 371350229 | 0,2945261 | 805 | 1 | 521117995 | 2,805084 |
| 5239 | 6 | 177143220 | 0,43461583 | 5239 | 6 | 177143220 | 0,2926575 | 3515 | 4 | 245947532 | 2,794010 |
| 2025 | 3 | 9124190 | 0,41976680 | 2025 | 3 | 9124190 | 0,2864646 | 3766 | 4 | 246990437 | 2,738281 |
| 1562 | 2 | 363489782 | 0,40269813 | 1562 | 2 | 363489782 | 0,2792628 | 3420 | 4 | 245552807 | 2,687064 |
| 5123 | 6 | 121839487 | 0,38105036 | 5123 | 6 | 121839487 | 0,2699920 | 5728 | 6 | 351298125 | 2,683066 |
| 3960 | 4 | 247796507 | 0,37141203 | 3960 | 4 | 247796507 | 0,2658120 | 5269 | 6 | 187827570 | 2,602368 |
| 3296 | 4 | 239237605 | 0,37106050 | 3296 | 4 | 239237605 | 0,2656589 | 1562 | 2 | 363489782 | 2,588188 |
| 3766 | 4 | 246990437 | 0,36777181 | 3766 | 4 | 246990437 | 0,2642244 | 2578 | 3 | 436987806 | 2,492446 |
| 3517 | 4 | 245955842 | 0,35240387 | 3517 | 4 | 245955842 | 0,2574669 | 1573 | 2 | 371350229 | 2,436171 |
| 3810 | 4 | 247173257 | 0,34103802 | 3810 | 4 | 247173257 | 0,2524090 | 3718 | 4 | 246790997 | 2,397791 |
| 4365 | 5 | 223840566 | 0,29382974 | 4365 | 5 | 223840566 | 0,2307876 | 3960 | 4 | 247796507 | 2,396755 |
| 5478 | 6 | 262261875 | 0,27894331 | 5478 | 6 | 262261875 | 0,2237387 | 2025 | 3 | 9124190 | 2,390213 |
| 4966 | 5 | 780442292 | 0,27664246 | 4966 | 5 | 780442292 | 0,2226382 | 1941 | 2 | 634317906 | 2,383527 |
| 5728 | 6 | 351298125 | 0,27652370 | 5728 | 6 | 351298125 | 0,2225813 | 4582 | 5 | 424809908 | 2,381976 |
| 3044 | 4 | 35384247 | 0,26648055 | 3044 | 4 | 35384247 | 0,2177406 | 1268 | 2 | 169989408 | 2,359644 |
| 4070 | 5 | 13981667 | 0,25348088 | 4070 | 5 | 13981667 | 0,2113829 | 5763 | 6 | 363763200 | 2,356424 |
| 382 | 1 | 262715518 | 0,25179534 | 382 | 1 | 262715518 | 0,2105506 | 5933 | 6 | 424307850 | 2,329228 |
| 3399 | 4 | 245465552 | 0,25044939 | 3399 | 4 | 245465552 | 0,2098846 | 4070 | 5 | 13981667 | 2,307127 |
| 1268 | 2 | 169989408 | 0,24212772 | 1268 | 2 | 169989408 | 0,2057396 | 5990 | 6 | 444608115 | 2,297095 |
| 5990 | 6 | 444608115 | 0,24192521 | 5990 | 6 | 444608115 | 0,2056381 | 842 | 1 | 545911919 | 2,284275 |
| 5034 | 6 | 78304583 | 0,23982158 | 5034 | 6 | 78304583 | 0,2045823 | 5954 | 6 | 431786895 | 2,269155 |
| 1617 | 2 | 402792016 | 0,22490490 | 1617 | 2 | 402792016 | 0,1970022 | 382 | 1 | 262715518 | 2,253707 |
| 2613 | 3 | 469546171 | 0,22229196 | 2613 | 3 | 469546171 | 0,1956567 | 5478 | 6 | 262261875 | 2,213185 |
| 842 | 1 | 545911919 | 0,21339410 | 842 | 1 | 545911919 | 0,1910332 | 5350 | 6 | 216675315 | 2,187097 |
| 4582 | 5 | 424809908 | 0,21246261 | 4582 | 5 | 424809908 | 0,1905453 | 2726 | 3 | 574663178 | 2,168420 |
| 5693 | 6 | 338833050 | 0,21225348 | 5693 | 6 | 338833050 | 0,1904357 | 5693 | 6 | 338833050 | 2,146788 |
| 3354 | 4 | 245278577 | 0,20578779 | 3354 | 4 | 245278577 | 0,1870271 | 3296 | 4 | 239237605 | 2,138550 |
| 745 | 1 | 480911633 | 0,20402532 | 745 | 1 | 480911633 | 0,1860915 | 4960 | 5 | 774885536 | 2,124039 |
| 3126 | 4 | 122795391 | 0,20287556 | 3126 | 4 | 122795391 | 0,1854796 | 3399 | 4 | 245465552 | 2,105284 |
| 794 | 1 | 513746829 | 0,20058913 | 794 | 1 | 513746829 | 0,1842591 | 2901 | 3 | 737455003 | 2,105209 |
| 5357 | 6 | 219168330 | 0,19933423 | 5357 | 6 | 219168330 | 0,1835872 | 4365 | 5 | 223840566 | 2,102395 |
| 4087 | 5 | 21770965 | 0,19553301 | 4087 | 5 | 21770965 | 0,1815426 | 4087 | 5 | 21770965 | 2,099218 |
| 251 | 1 | 183647544 | 0,19490304 | 251 | 1 | 183647544 | 0,1812024 | 933 | 1 | 606891568 | 2,090111 |
| 4682 | 5 | 517422508 | 0,19407516 | 4682 | 5 | 517422508 | 0,1807547 | 3044 | 4 | 35384247 | 2,088211 |
| 4795 | 5 | 622074746 | 0,19067677 | 4795 | 5 | 622074746 | 0,1789099 | 620 | 1 | 397148378 | 2,085016 |
| 5175 | 6 | 151958886 | 0,17890416 | 5175 | 6 | 151958886 | 0,1724261 | 4838 | 5 | 661898164 | 2,083066 |
| 2101 | 3 | 110715956 | 0,17504233 | 2101 | 3 | 110715956 | 0,1702661 | 5175 | 6 | 151958886 | 2,081309 |
| 1657 | 2 | 431375459 | 0,17330293 | 1657 | 2 | 431375459 | 0,1692876 | 2480 | 3 | 345824384 | 2,068159 |
| 2480 | 3 | 345824384 | 0,16749614 | 2480 | 3 | 345824384 | 0,1659949 | 3532 | 4 | 246018167 | 2,065721 |
| 3658 | 4 | 246541697 | 0,16629752 | 3658 | 4 | 246541697 | 0,1653101 | 5789 | 6 | 373022970 | 2,064492 |
| 1402 | 2 | 249156009 | 0,16352084 | 1402 | 2 | 249156009 | 0,1637168 | 2584 | 3 | 442569240 | 2,058015 |
| 3963 | 4 | 247808972 | 0,16303134 | 3963 | 4 | 247808972 | 0,1634349 | 3881 | 4 | 247468262 | 2,057619 |
| 3259 | 4 | 210236675 | 0,16088305 | 3259 | 4 | 210236675 | 0,1621940 | 5756 | 6 | 361270185 | 2,037824 |
| 5789 | 6 | 373022970 | 0,15911407 | 5789 | 6 | 373022970 | 0,1611676 | 1657 | 2 | 431375459 | 2,033325 |
| 1113 | 2 | 24893463 | 0,15303764 | 1113 | 2 | 24893463 | 0,1576097 | 4604 | 5 | 445184680 | 2,018235 |
| 3964 | 4 | 247813127 | 0,15117487 | 3964 | 4 | 247813127 | 0,1565086 | 2968 | 3 | 799781016 | 2,013910 |
| 1865 | 2 | 580009364 | 0,15102129 | 1865 | 2 | 580009364 | 0,1564176 | 2221 | 3 | 178234880 | 2,001622 |

Fonte: Da autora (2019)

Tabela 36 – Bayes A modificado \times GWAS
cenário Poligênico, $h^2 = 0,5$

| Bayes A modificado | | | | | | | | GWAS | | | |
|--------------------|----|-----------|-------------|-------|----|-----------|--------------------------------|-------|----|-----------|--------------------------------|
| Marca | Cr | Pos | wald | Marca | Cr | Pos | $-\log_{10}(\text{valor} - p)$ | Marca | Cr | Pos | $-\log_{10}(\text{valor} - p)$ |
| 5826 | 6 | 386200335 | 41,54988607 | 5826 | 6 | 386200335 | 9,9396755 | 5826 | 6 | 386200335 | 8,459814 |
| 3480 | 4 | 245802107 | 26,26324405 | 3480 | 4 | 245802107 | 6,5259225 | 3480 | 4 | 245802107 | 6,266500 |
| 1562 | 2 | 363489782 | 25,27598478 | 1562 | 2 | 363489782 | 6,3037689 | 1562 | 2 | 363489782 | 5,390297 |
| 4317 | 5 | 204516414 | 6,49954685 | 4317 | 5 | 204516414 | 1,9669705 | 3261 | 4 | 212893509 | 4,081755 |
| 3261 | 4 | 212893509 | 6,19393799 | 3261 | 4 | 212893509 | 1,8921505 | 5673 | 6 | 331710150 | 3,983664 |
| 1093 | 2 | 20954874 | 4,42335303 | 1093 | 2 | 20954874 | 1,4503800 | 3517 | 4 | 245955842 | 3,790299 |
| 2517 | 3 | 380243227 | 3,68683255 | 2517 | 3 | 380243227 | 1,2608735 | 5846 | 6 | 393323235 | 3,772569 |
| 3517 | 4 | 245955842 | 3,66122520 | 3517 | 4 | 245955842 | 1,2542020 | 2517 | 3 | 380243227 | 3,400196 |
| 4337 | 5 | 209498142 | 1,21791431 | 4337 | 5 | 209498142 | 0,5690058 | 5933 | 6 | 424307850 | 3,338913 |
| 4629 | 5 | 468337830 | 1,16161645 | 4629 | 5 | 468337830 | 0,5510920 | 5277 | 6 | 190676730 | 3,331761 |
| 5673 | 6 | 331710150 | 1,13575446 | 5673 | 6 | 331710150 | 0,5427977 | 1646 | 2 | 423515013 | 3,277031 |
| 499 | 1 | 291031959 | 0,85380575 | 499 | 1 | 291031959 | 0,4491874 | 3736 | 4 | 246865787 | 3,273612 |
| 4913 | 5 | 731357614 | 0,73313200 | 4913 | 5 | 731357614 | 0,4068581 | 5884 | 6 | 406856745 | 3,248133 |
| 3971 | 4 | 247842212 | 0,67967793 | 3971 | 4 | 247842212 | 0,3875365 | 4317 | 5 | 204516414 | 3,192595 |
| 4987 | 5 | 799890938 | 0,63280172 | 4987 | 5 | 799890938 | 0,3702546 | 4649 | 5 | 486860350 | 3,177421 |
| 1417 | 2 | 259874800 | 0,62634284 | 1417 | 2 | 259874800 | 0,3678466 | 2749 | 3 | 596058675 | 3,177365 |
| 945 | 1 | 614932841 | 0,52636941 | 945 | 1 | 614932841 | 0,3296268 | 5862 | 6 | 399021555 | 3,117959 |
| 5461 | 6 | 256207410 | 0,52189416 | 5461 | 6 | 256207410 | 0,3278700 | 3381 | 4 | 245390762 | 3,105844 |
| 4549 | 5 | 394247750 | 0,47953715 | 4549 | 5 | 394247750 | 0,3110181 | 5670 | 6 | 330641715 | 3,008389 |
| 5277 | 6 | 190676730 | 0,46728833 | 5277 | 6 | 190676730 | 0,3060645 | 187 | 1 | 83595286 | 2,927185 |
| 5577 | 6 | 297520230 | 0,45589961 | 5577 | 6 | 297520230 | 0,3014240 | 5799 | 6 | 376584420 | 2,906733 |
| 750 | 1 | 484262163 | 0,44488001 | 750 | 1 | 484262163 | 0,2969009 | 1093 | 2 | 20954874 | 2,879809 |
| 3005 | 4 | 2637306 | 0,42634915 | 3005 | 4 | 2637306 | 0,2892179 | 1865 | 2 | 580009364 | 2,858210 |
| 1263 | 2 | 169265278 | 0,41477034 | 1263 | 2 | 169265278 | 0,2843658 | 3389 | 4 | 245424002 | 2,708215 |
| 3960 | 4 | 247796507 | 0,41156050 | 3960 | 4 | 247796507 | 0,2830135 | 5033 | 6 | 75187712 | 2,700129 |
| 1295 | 2 | 186512163 | 0,40301215 | 1295 | 2 | 186512163 | 0,2793961 | 5805 | 6 | 378721290 | 2,686021 |
| 4453 | 5 | 305339654 | 0,37175171 | 4453 | 5 | 305339654 | 0,2659599 | 427 | 1 | 268373293 | 2,590842 |
| 1354 | 2 | 214263400 | 0,34976706 | 1354 | 2 | 214263400 | 0,2562982 | 4290 | 5 | 193521948 | 2,568202 |
| 514 | 1 | 293632633 | 0,34727277 | 514 | 1 | 293632633 | 0,2551901 | 945 | 1 | 614932841 | 2,566545 |
| 187 | 1 | 83595286 | 0,33824779 | 187 | 1 | 83595286 | 0,2511592 | 5698 | 6 | 340613775 | 2,563206 |
| 2346 | 3 | 223513366 | 0,33666996 | 2346 | 3 | 223513366 | 0,2504509 | 4265 | 5 | 184588958 | 2,491110 |
| 1534 | 2 | 343481372 | 0,33335430 | 1534 | 2 | 343481372 | 0,2489592 | 4838 | 5 | 661898164 | 2,438240 |
| 4649 | 5 | 486860350 | 0,31810908 | 4649 | 5 | 486860350 | 0,2420378 | 4453 | 5 | 305339654 | 2,422658 |
| 4170 | 5 | 75234294 | 0,29086884 | 4170 | 5 | 75234294 | 0,2293952 | 1095 | 2 | 209583707 | 2,421069 |
| 4205 | 5 | 145121333 | 0,28942461 | 4205 | 5 | 145121333 | 0,2287143 | 3133 | 4 | 128632208 | 2,401665 |
| 4497 | 5 | 346089198 | 0,27517380 | 4497 | 5 | 346089198 | 0,2219341 | 3096 | 4 | 76123039 | 2,394804 |
| 1252 | 2 | 164561951 | 0,27107091 | 1252 | 2 | 164561951 | 0,2199606 | 5619 | 6 | 312478320 | 2,369837 |
| 162 | 1 | 56885616 | 0,27036730 | 162 | 1 | 56885616 | 0,2196211 | 2877 | 3 | 715129267 | 2,355939 |
| 5365 | 6 | 222017490 | 0,27003711 | 5365 | 6 | 222017490 | 0,2194617 | 3971 | 4 | 247842212 | 2,332747 |
| 898 | 1 | 583437857 | 0,26339135 | 898 | 1 | 583437857 | 0,2162395 | 4802 | 5 | 628557628 | 2,332568 |
| 2749 | 3 | 596058675 | 0,26150224 | 2749 | 3 | 596058675 | 0,2153186 | 4648 | 5 | 485934224 | 2,326017 |
| 1230 | 2 | 144528591 | 0,25310874 | 1230 | 2 | 144528591 | 0,2111993 | 5244 | 6 | 178923945 | 2,324638 |
| 1095 | 2 | 20958307 | 0,24937258 | 1095 | 2 | 20958307 | 0,2093509 | 1358 | 2 | 214423504 | 2,304391 |
| 1319 | 2 | 198756269 | 0,24683743 | 1319 | 2 | 198756269 | 0,2080913 | 3972 | 4 | 247846367 | 2,286949 |
| 2365 | 3 | 238846899 | 0,24485582 | 2365 | 3 | 238846899 | 0,2071037 | 499 | 1 | 291031959 | 2,284529 |
| 1210 | 2 | 128180134 | 0,24025469 | 1210 | 2 | 128180134 | 0,2047999 | 1185 | 2 | 58354695 | 2,283718 |
| 4266 | 5 | 185596985 | 0,23445930 | 4266 | 5 | 185596985 | 0,2018766 | 5103 | 6 | 113384656 | 2,265262 |
| 821 | 1 | 531839692 | 0,23303159 | 821 | 1 | 531839692 | 0,2011526 | 1321 | 2 | 198757428 | 2,264077 |
| 4081 | 5 | 19973330 | 0,22614574 | 4081 | 5 | 19973330 | 0,1976392 | 821 | 1 | 531839692 | 2,263088 |
| 2936 | 3 | 770013368 | 0,22104651 | 2936 | 3 | 770013368 | 0,1950135 | 308 | 1 | 216989434 | 2,252107 |
| 1243 | 2 | 159619959 | 0,22093496 | 1243 | 2 | 159619959 | 0,1949558 | 1354 | 2 | 214263400 | 2,231001 |
| 3389 | 4 | 245424002 | 0,21836961 | 3389 | 4 | 245424002 | 0,1936267 | 5547 | 6 | 286835880 | 2,222143 |
| 4838 | 5 | 661898164 | 0,21706523 | 4838 | 5 | 661898164 | 0,1929488 | 4629 | 5 | 468337830 | 2,186507 |
| 2746 | 3 | 593267958 | 0,21489026 | 2746 | 3 | 593267958 | 0,1918153 | 4170 | 5 | 75234294 | 2,178396 |
| 1358 | 2 | 214423504 | 0,20798527 | 1358 | 2 | 214423504 | 0,1881897 | 3992 | 4 | 247929467 | 2,174125 |
| 5962 | 6 | 434636055 | 0,20572016 | 5962 | 6 | 434636055 | 0,1869913 | 3005 | 4 | 2637306 | 2,168391 |
| 4969 | 5 | 783220670 | 0,20419337 | 4969 | 5 | 783220670 | 0,1861808 | 2072 | 3 | 53549251 | 2,162393 |
| 3918 | 4 | 247621997 | 0,19909875 | 3918 | 4 | 247621997 | 0,1834609 | 1417 | 2 | 259874800 | 2,154835 |
| 182 | 1 | 74579886 | 0,18789776 | 182 | 1 | 74579886 | 0,1773926 | 5789 | 6 | 373022970 | 2,123864 |
| 970 | 1 | 631685492 | 0,18782542 | 970 | 1 | 631685492 | 0,1773530 | 1561 | 2 | 362775196 | 2,111705 |

Fonte: Da autora (2019)

Tabela 37 – Bayes A modificado \times GWAS
cenário Poligênico, $h^2 = 0,8$

| Bayes A modificado | | | | | | | | GWAS | | | |
|--------------------|----|-----------|------------|-------|----|-----------|--------------------------------|-------|----|-----------|--------------------------------|
| Marca | Cr | Pos | wald | Marca | Cr | Pos | $-\log_{10}(\text{valor} - p)$ | Marca | Cr | Pos | $-\log_{10}(\text{valor} - p)$ |
| 5826 | 6 | 386200335 | 73,9293430 | 5826 | 6 | 386200335 | 17,0917091 | 5826 | 6 | 386200335 | 11,587916 |
| 1562 | 2 | 363489782 | 46,8210410 | 1562 | 2 | 363489782 | 11,1091662 | 3480 | 4 | 245802107 | 8,318912 |
| 3480 | 4 | 245802107 | 45,5758688 | 3480 | 4 | 245802107 | 10,8331571 | 5954 | 6 | 431786895 | 5,722483 |
| 2517 | 3 | 380243227 | 17,3014008 | 2517 | 3 | 380243227 | 4,4962751 | 1562 | 2 | 363489782 | 5,662491 |
| 1093 | 2 | 20954874 | 12,0016842 | 1093 | 2 | 20954874 | 3,2744763 | 5033 | 6 | 75187712 | 5,350154 |
| 4549 | 5 | 394247750 | 10,6081518 | 4549 | 5 | 394247750 | 2,9484995 | 5862 | 6 | 399021555 | 4,623745 |
| 3517 | 4 | 245955842 | 9,2493658 | 3517 | 4 | 245955842 | 2,6278671 | 2517 | 3 | 380243227 | 4,065757 |
| 3960 | 4 | 247796507 | 8,4597624 | 3960 | 4 | 247796507 | 2,4399866 | 5590 | 6 | 302150115 | 3,514904 |
| 3261 | 4 | 212893509 | 6,6959659 | 3261 | 4 | 212893509 | 2,0148822 | 3736 | 4 | 246865787 | 3,498160 |
| 4629 | 5 | 468337830 | 4,4155701 | 4629 | 5 | 468337830 | 1,4483990 | 1093 | 2 | 20954874 | 3,365941 |
| 3288 | 4 | 235779015 | 4,2546090 | 3288 | 4 | 235779015 | 1,4073353 | 2292 | 3 | 200201258 | 3,072857 |
| 5862 | 6 | 399021555 | 2,9436847 | 5862 | 6 | 399021555 | 1,0644229 | 3517 | 4 | 245955842 | 3,015800 |
| 1528 | 2 | 339193855 | 1,9590123 | 1528 | 2 | 339193855 | 0,7915076 | 859 | 1 | 557303721 | 2,967605 |
| 5033 | 6 | 75187712 | 1,9201536 | 5033 | 6 | 75187712 | 0,7803114 | 3643 | 4 | 246479372 | 2,908547 |
| 5560 | 6 | 291465765 | 1,5082591 | 5560 | 6 | 291465765 | 0,6587536 | 3996 | 4 | 247946087 | 2,879696 |
| 3476 | 4 | 245785487 | 1,1834999 | 3476 | 4 | 245785487 | 0,5580778 | 89 | 1 | 28084996 | 2,879099 |
| 5636 | 6 | 318532785 | 1,1765505 | 5636 | 6 | 318532785 | 0,5558625 | 3918 | 4 | 247621997 | 2,809284 |
| 382 | 1 | 262715518 | 0,9746634 | 382 | 1 | 262715518 | 0,4900990 | 5134 | 6 | 131576099 | 2,790126 |
| 2776 | 3 | 621175128 | 0,8263504 | 2776 | 3 | 621175128 | 0,4396988 | 1865 | 2 | 580009364 | 2,786465 |
| 3632 | 4 | 246433667 | 0,8023517 | 3632 | 4 | 246433667 | 0,4313394 | 4629 | 5 | 468337830 | 2,781031 |
| 4085 | 5 | 20200987 | 0,7979448 | 4085 | 5 | 20200987 | 0,4297974 | 5636 | 6 | 318532785 | 2,764711 |
| 5590 | 6 | 302150115 | 0,7692613 | 5590 | 6 | 302150115 | 0,4197074 | 1304 | 2 | 189293647 | 2,700383 |
| 5643 | 6 | 321025800 | 0,7246898 | 5643 | 6 | 321025800 | 0,4038321 | 5186 | 6 | 157129197 | 2,698615 |
| 3406 | 4 | 245494637 | 0,7065538 | 3406 | 4 | 245494637 | 0,3972996 | 5063 | 6 | 83064740 | 2,698013 |
| 4895 | 5 | 714687346 | 0,7062301 | 4895 | 5 | 714687346 | 0,3971826 | 5805 | 6 | 378721290 | 2,687730 |
| 4341 | 5 | 210890745 | 0,6541248 | 4341 | 5 | 210890745 | 0,3781574 | 3476 | 4 | 245785487 | 2,630147 |
| 5131 | 6 | 129927781 | 0,6107047 | 5131 | 6 | 129927781 | 0,3619876 | 1646 | 2 | 423515013 | 2,593889 |
| 89 | 1 | 28084996 | 0,5996509 | 89 | 1 | 28084996 | 0,3578212 | 5801 | 6 | 377296710 | 2,582850 |
| 3918 | 4 | 247621997 | 0,5975740 | 3918 | 4 | 247621997 | 0,3570360 | 2840 | 3 | 680710424 | 2,569747 |
| 3032 | 4 | 28192994 | 0,5868215 | 3032 | 4 | 28192994 | 0,3529587 | 966 | 1 | 629005068 | 2,547537 |
| 3170 | 4 | 153293452 | 0,5671063 | 3170 | 4 | 153293452 | 0,3454278 | 5530 | 6 | 280781415 | 2,532024 |
| 5109 | 6 | 113385418 | 0,5294514 | 5109 | 6 | 113385418 | 0,3308342 | 4459 | 5 | 310896410 | 2,531509 |
| 4849 | 5 | 672085550 | 0,5080592 | 4849 | 5 | 672085550 | 0,3224111 | 5416 | 6 | 240180885 | 2,488117 |
| 5416 | 6 | 240180885 | 0,5066268 | 5416 | 6 | 240180885 | 0,3218435 | 1933 | 2 | 628601218 | 2,436105 |
| 859 | 1 | 557303721 | 0,4866212 | 859 | 1 | 557303721 | 0,3138659 | 3658 | 4 | 246541697 | 2,433635 |
| 2736 | 3 | 583965568 | 0,4745069 | 2736 | 3 | 583965568 | 0,3089883 | 3538 | 4 | 246043097 | 2,428786 |
| 1646 | 2 | 423515013 | 0,4689426 | 1646 | 2 | 423515013 | 0,3067357 | 5442 | 6 | 249440655 | 2,404876 |
| 3318 | 4 | 245131700 | 0,4530182 | 3318 | 4 | 245131700 | 0,3002445 | 3220 | 4 | 180309373 | 2,391637 |
| 3389 | 4 | 245424002 | 0,4434109 | 3389 | 4 | 245424002 | 0,2962954 | 5290 | 6 | 195306615 | 2,384984 |
| 3363 | 4 | 245315972 | 0,4329097 | 3363 | 4 | 245315972 | 0,2919492 | 382 | 1 | 262715518 | 2,379969 |
| 4459 | 5 | 310896410 | 0,4304991 | 4459 | 5 | 310896410 | 0,2909471 | 5782 | 6 | 370529955 | 2,373650 |
| 522 | 1 | 294646235 | 0,4140458 | 522 | 1 | 294646235 | 0,2840609 | 1086 | 2 | 19837545 | 2,371760 |
| 5472 | 6 | 260125005 | 0,4102730 | 5472 | 6 | 260125005 | 0,2824701 | 4549 | 5 | 394247750 | 2,352150 |
| 3160 | 4 | 148408488 | 0,4016882 | 3160 | 4 | 148408488 | 0,2788337 | 2998 | 3 | 827688186 | 2,352145 |
| 4650 | 5 | 487786476 | 0,3941571 | 4650 | 5 | 487786476 | 0,2756239 | 1880 | 2 | 590728155 | 2,342969 |
| 2279 | 3 | 194643731 | 0,3695697 | 2279 | 3 | 194643731 | 0,2650091 | 5698 | 6 | 340613775 | 2,338420 |
| 4648 | 5 | 485934224 | 0,3612541 | 4648 | 5 | 485934224 | 0,2613696 | 3295 | 4 | 238553534 | 2,331521 |
| 1902 | 2 | 606449049 | 0,3575713 | 1902 | 2 | 606449049 | 0,2597493 | 5171 | 6 | 149251173 | 2,318388 |
| 4698 | 5 | 532240524 | 0,3558918 | 4698 | 5 | 532240524 | 0,2590086 | 3337 | 4 | 245207942 | 2,312483 |
| 3220 | 4 | 180309373 | 0,3512323 | 3220 | 4 | 180309373 | 0,2569480 | 728 | 1 | 469519830 | 2,297564 |
| 3448 | 4 | 245669147 | 0,3415666 | 3448 | 4 | 245669147 | 0,2526454 | 1686 | 2 | 452098456 | 2,294555 |
| 3755 | 4 | 246944732 | 0,3341853 | 3755 | 4 | 246944732 | 0,2493335 | 3632 | 4 | 246433667 | 2,263630 |
| 3845 | 4 | 247318682 | 0,3339246 | 3845 | 4 | 247318682 | 0,2492161 | 3261 | 4 | 212893509 | 2,223022 |
| 5145 | 6 | 137326250 | 0,3281173 | 5145 | 6 | 137326250 | 0,2465933 | 5641 | 6 | 320313510 | 2,207166 |
| 4940 | 5 | 756363016 | 0,3267869 | 4940 | 5 | 756363016 | 0,2459903 | 5244 | 6 | 178923945 | 2,205616 |
| 1630 | 2 | 412081635 | 0,3247530 | 1630 | 2 | 412081635 | 0,2450670 | 4849 | 5 | 672085550 | 2,204546 |
| 5829 | 6 | 387268770 | 0,3218578 | 5829 | 6 | 387268770 | 0,2437494 | 1249 | 2 | 161879250 | 2,186632 |
| 158 | 1 | 56175806 | 0,3118090 | 158 | 1 | 56175806 | 0,2391463 | 5336 | 6 | 211689285 | 2,170129 |
| 1686 | 2 | 452098456 | 0,3056386 | 1686 | 2 | 452098456 | 0,2362959 | 4394 | 5 | 250698220 | 2,163688 |
| 3382 | 4 | 245394917 | 0,3037942 | 3382 | 4 | 245394917 | 0,2354402 | 1867 | 2 | 581438536 | 2,150772 |

Fonte: Da autora (2019)

Tabelas Bayes B modificado x GWAS

Tabela 38 – Bayes B modificado × GWAS
cenário Oligogênico, $h^2 = 0,2$

| Bayes B modificado | | | | | | | | GWAS | | | |
|--------------------|----|-----------|-------------|-------|----|-----------|--------------------------------|-------|----|-----------|--------------------------------|
| Marca | Cr | Pos | wald | Marca | Cr | Pos | $-\log_{10}(\text{valor} - p)$ | Marca | Cr | Pos | $-\log_{10}(\text{valor} - p)$ |
| 4004 | 5 | 2123914 | 27,83047608 | 4004 | 5 | 2123914 | 6,8780345 | 4004 | 5 | 2123914 | 7,597231 |
| 2983 | 3 | 813734601 | 11,06416008 | 2983 | 3 | 813734601 | 3,0554559 | 2983 | 3 | 813734601 | 4,887828 |
| 1826 | 2 | 552140507 | 2,19479803 | 1826 | 2 | 552140507 | 0,8586212 | 4260 | 5 | 181672656 | 4,265723 |
| 4877 | 5 | 698017078 | 1,22546070 | 4877 | 5 | 698017078 | 0,5713929 | 1029 | 2 | 4679573 | 3,633500 |
| 1029 | 2 | 4679573 | 0,98203897 | 1029 | 2 | 4679573 | 0,4925544 | 2896 | 3 | 732803808 | 3,457414 |
| 279 | 1 | 202620610 | 0,74714602 | 279 | 1 | 202620610 | 0,4118613 | 4303 | 5 | 200813183 | 3,316936 |
| 4045 | 5 | 9279776 | 0,72897568 | 4045 | 5 | 9279776 | 0,4053695 | 4877 | 5 | 698017078 | 3,296434 |
| 2897 | 3 | 733734047 | 0,69463067 | 2897 | 3 | 733734047 | 0,3929808 | 2897 | 3 | 733734047 | 3,229439 |
| 2256 | 3 | 186580149 | 0,65084726 | 2256 | 3 | 186580149 | 0,3769473 | 4047 | 5 | 9965534 | 3,221998 |
| 1051 | 2 | 7482709 | 0,60792093 | 1051 | 2 | 7482709 | 0,3609404 | 4621 | 5 | 460928822 | 3,214396 |

Fonte: Da autora (2019)

Tabela 39 – Bayes B modificado × GWAS
cenário Oligogênico, $h^2 = 0,5$

| Bayes B modificado | | | | | | | | GWAS | | | |
|--------------------|----|-----------|-------------|-------|----|-----------|--------------------------------|-------|----|-----------|--------------------------------|
| Marca | Cr | Pos | wald | Marca | Cr | Pos | $-\log_{10}(\text{valor} - p)$ | Marca | Cr | Pos | $-\log_{10}(\text{valor} - p)$ |
| 4004 | 5 | 2123914 | 68,12398818 | 4004 | 5 | 2123914 | 15,8138025 | 4004 | 5 | 2123914 | 13,339721 |
| 3677 | 4 | 246620642 | 56,20035255 | 3677 | 4 | 246620642 | 13,1840917 | 3677 | 4 | 246620642 | 9,102989 |
| 4877 | 5 | 698017078 | 47,62465989 | 4877 | 5 | 698017078 | 11,2872231 | 4877 | 5 | 698017078 | 7,418931 |
| 193 | 1 | 88380847 | 3,34000265 | 193 | 1 | 88380847 | 11,699602 | 3805 | 4 | 247152482 | 5,668896 |
| 538 | 1 | 298413091 | 1,12026531 | 538 | 1 | 298413091 | 0,5378097 | 4260 | 5 | 181672656 | 5,507352 |
| 713 | 1 | 459468239 | 0,90587606 | 713 | 1 | 459468239 | 0,4669770 | 4625 | 5 | 464633326 | 5,215285 |
| 532 | 1 | 297315711 | 0,63878172 | 532 | 1 | 297315711 | 0,3724782 | 3933 | 4 | 247684322 | 4,264209 |
| 547 | 1 | 348230637 | 0,63788754 | 547 | 1 | 348230637 | 0,3721461 | 3549 | 4 | 246088802 | 3,997775 |
| 186 | 1 | 82563962 | 0,57118260 | 186 | 1 | 82563962 | 0,3469908 | 4772 | 5 | 600773848 | 3,850929 |
| 309 | 1 | 216989646 | 0,56023028 | 309 | 1 | 216989646 | 0,3427840 | 3037 | 4 | 29209539 | 3,656472 |

Fonte: Da autora (2019)

Tabela 40 – Bayes B modificado × GWAS
cenário Oligogênico, $h^2 = 0,8$

| Bayes B modificado | | | | | | | | GWAS | | | |
|--------------------|----|-----------|--------------|-------|----|-----------|--------------------------------|-------|----|-----------|--------------------------------|
| Marca | Cr | Pos | wald | Marca | Cr | Pos | $-\log_{10}(\text{valor} - p)$ | Marca | Cr | Pos | $-\log_{10}(\text{valor} - p)$ |
| 4004 | 5 | 2123914 | 270,40218712 | 4004 | 5 | 2123914 | 60,03274544 | 4004 | 5 | 2123914 | 38,278105 |
| 4877 | 5 | 698017078 | 130,32772799 | 4877 | 5 | 698017078 | 29,45915560 | 4877 | 5 | 698017078 | 15,782135 |
| 3677 | 4 | 246620642 | 67,90243253 | 3677 | 4 | 246620642 | 15,76500438 | 4260 | 5 | 181672656 | 10,589633 |
| 538 | 1 | 298413091 | 17,26,60358 | 538 | 1 | 298413091 | 4,48717910 | 3677 | 4 | 246620642 | 6,920932 |
| 2088 | 3 | 69892547 | 1,217100162 | 2088 | 3 | 69892547 | 3,31391602 | 4303 | 5 | 200813183 | 6,386065 |
| 2000 | 2 | 676478485 | 1,30967946 | 2000 | 2 | 676478485 | 0,59781998 | 4237 | 5 | 171688991 | 5,372977 |
| 900 | 1 | 584778069 | 0,46368477 | 900 | 1 | 584778069 | 0,30459984 | 5412 | 6 | 238756305 | 4,982878 |
| 4927 | 5 | 744323378 | 0,42962634 | 4927 | 5 | 744323378 | 0,29058387 | 4625 | 5 | 464633326 | 4,885838 |
| 189 | 1 | 87910951 | 0,37879464 | 189 | 1 | 87910951 | 0,26901674 | 4047 | 5 | 9965534 | 4,743689 |
| 5764 | 6 | 364119345 | 0,37609188 | 5764 | 6 | 364119345 | 0,26784577 | 4772 | 5 | 600773848 | 3,910513 |

Fonte: Da autora (2019)

Tabela 41 – Bayes B modificado \times GWAS
cenário Poligênico, $h^2 = 0,2$

| Bayes B modificado | | | | | | | GWAS | | | | |
|--------------------|----|-----------|------------|-------|----|-----------|--------------------------------|-------|----|-----------|--------------------------------|
| Marca | Cr | Pos | wald | Marca | Cr | Pos | $-\log_{10}(\text{valor} - p)$ | Marca | Cr | Pos | $-\log_{10}(\text{valor} - p)$ |
| 3480 | 4 | 245802107 | 6,20564727 | 3480 | 4 | 245802107 | 1,8950236 | 5826 | 6 | 386200335 | 4,358662 |
| 5826 | 6 | 386200335 | 3,89694552 | 5826 | 6 | 386200335 | 1,3153882 | 3480 | 4 | 245802107 | 4,130539 |
| 3941 | 4 | 247717562 | 2,78314167 | 3941 | 4 | 247717562 | 1,0210842 | 3941 | 4 | 247717562 | 3,956526 |
| 5862 | 6 | 399021555 | 2,40779958 | 5862 | 6 | 399021555 | 0,9181777 | 5862 | 6 | 399021555 | 3,582077 |
| 4966 | 5 | 780442292 | 1,78525023 | 4966 | 5 | 780442292 | 0,7411091 | 4966 | 5 | 780442292 | 3,359942 |
| 5239 | 6 | 177143220 | 1,51739910 | 5239 | 6 | 177143220 | 0,6615166 | 5801 | 6 | 377296710 | 3,308852 |
| 3960 | 4 | 247796507 | 1,50939889 | 3960 | 4 | 247796507 | 0,6590984 | 2840 | 3 | 680710424 | 3,179535 |
| 5801 | 6 | 377296710 | 1,41573819 | 5801 | 6 | 377296710 | 0,6305872 | 5239 | 6 | 177143220 | 3,076366 |
| 3718 | 4 | 246790997 | 1,08834622 | 3718 | 4 | 246790997 | 0,5274806 | 3517 | 4 | 245955842 | 3,035031 |
| 1562 | 2 | 363489782 | 0,78563287 | 1562 | 2 | 363489782 | 0,4254781 | 1408 | 2 | 253443525 | 2,992735 |
| 1573 | 2 | 371350229 | 0,70839472 | 1573 | 2 | 371350229 | 0,3979647 | 5123 | 6 | 121839487 | 2,850738 |
| 5990 | 6 | 444608115 | 0,61391934 | 5990 | 6 | 444608115 | 0,3631954 | 3658 | 4 | 246541697 | 2,838595 |
| 2578 | 3 | 436987806 | 0,55374066 | 2578 | 3 | 436987806 | 0,3402803 | 3810 | 4 | 247173257 | 2,817565 |
| 3810 | 4 | 247173257 | 0,48462915 | 3810 | 4 | 247173257 | 0,3130663 | 805 | 1 | 521117995 | 2,805084 |
| 805 | 1 | 521117995 | 0,47570530 | 805 | 1 | 521117995 | 0,3094725 | 3515 | 4 | 245947532 | 2,764010 |
| 620 | 1 | 397148378 | 0,47317843 | 620 | 1 | 397148378 | 0,3084512 | 3766 | 4 | 246990437 | 2,738281 |
| 2840 | 3 | 680710424 | 0,46984749 | 2840 | 3 | 680710424 | 0,3071026 | 3420 | 4 | 245552807 | 2,687064 |
| 3515 | 4 | 245947532 | 0,39830694 | 3515 | 4 | 245947532 | 0,2773949 | 5728 | 6 | 351298125 | 2,683066 |
| 4365 | 5 | 223840566 | 0,38630090 | 4365 | 5 | 223840566 | 0,2722553 | 5269 | 6 | 187827570 | 2,602368 |
| 5123 | 6 | 121839487 | 0,33426440 | 5123 | 6 | 121839487 | 0,2493691 | 1562 | 2 | 363489782 | 2,588188 |
| 3961 | 4 | 247800662 | 0,32065249 | 3961 | 4 | 247800662 | 0,2431998 | 2578 | 3 | 436987806 | 2,492446 |
| 3399 | 4 | 245465552 | 0,31363366 | 3399 | 4 | 245465552 | 0,2399857 | 1573 | 2 | 371350229 | 2,436171 |
| 2901 | 3 | 737455003 | 0,31242030 | 2901 | 3 | 737455003 | 0,2394277 | 3718 | 4 | 246790997 | 2,397791 |
| 842 | 1 | 545911919 | 0,31189515 | 842 | 1 | 545911919 | 0,2391859 | 3960 | 4 | 247796507 | 2,396755 |
| 5350 | 6 | 216675315 | 0,30682783 | 5350 | 6 | 216675315 | 0,2368467 | 2025 | 3 | 9124190 | 2,390213 |
| 1955 | 2 | 644322111 | 0,29445531 | 1955 | 2 | 644322111 | 0,2310812 | 1941 | 2 | 634317906 | 2,383527 |
| 4070 | 5 | 13981667 | 0,29041941 | 4070 | 5 | 13981667 | 0,2291834 | 4582 | 5 | 424809908 | 2,381976 |
| 3296 | 4 | 239237605 | 0,27875281 | 3296 | 4 | 239237605 | 0,2236477 | 1268 | 2 | 169989408 | 2,359644 |
| 5756 | 6 | 361270185 | 0,27846822 | 5756 | 6 | 361270185 | 0,2235117 | 5763 | 6 | 363763200 | 2,356424 |
| 3532 | 4 | 246018167 | 0,27684936 | 3532 | 4 | 246018167 | 0,2227373 | 5933 | 6 | 424307850 | 2,329228 |
| 1408 | 2 | 253443525 | 0,26611608 | 1408 | 2 | 253443525 | 0,2175638 | 4070 | 5 | 13981667 | 2,307127 |
| 933 | 1 | 606891568 | 0,26544691 | 933 | 1 | 606891568 | 0,2172390 | 5990 | 6 | 444608115 | 2,297095 |
| 5728 | 6 | 351298125 | 0,26347931 | 5728 | 6 | 351298125 | 0,2162823 | 842 | 1 | 545911919 | 2,284275 |
| 4960 | 5 | 774885536 | 0,22509819 | 4960 | 5 | 774885536 | 0,1971015 | 5954 | 6 | 431786895 | 2,269155 |
| 656 | 1 | 421272195 | 0,21243156 | 656 | 1 | 421272195 | 0,1905290 | 382 | 1 | 262715518 | 2,253707 |
| 3940 | 4 | 247713407 | 0,21242672 | 3940 | 4 | 247713407 | 0,1905265 | 5478 | 6 | 262261875 | 2,213185 |
| 3104 | 4 | 90203655 | 0,20139205 | 3104 | 4 | 90203655 | 0,1846882 | 5350 | 6 | 216675315 | 2,187097 |
| 2998 | 3 | 827688186 | 0,20042771 | 2998 | 3 | 827688186 | 0,1841728 | 2726 | 3 | 574663178 | 2,168420 |
| 2726 | 3 | 574663178 | 0,19544283 | 2726 | 3 | 574663178 | 0,1814939 | 5693 | 6 | 338833050 | 2,146788 |
| 3420 | 4 | 245552807 | 0,19091776 | 3420 | 4 | 245552807 | 0,1790411 | 3296 | 4 | 239237605 | 2,138550 |
| 3517 | 4 | 245955842 | 0,19047893 | 3517 | 4 | 245955842 | 0,1788021 | 4960 | 5 | 774885536 | 2,124039 |
| 1776 | 2 | 516411203 | 0,18467615 | 1776 | 2 | 516411203 | 0,1756235 | 3399 | 4 | 245465552 | 2,105284 |
| 3680 | 4 | 246633107 | 0,18449970 | 3680 | 4 | 246633107 | 0,1755263 | 2901 | 3 | 737455003 | 2,105209 |
| 3044 | 4 | 35384247 | 0,18289168 | 3044 | 4 | 35384247 | 0,1746389 | 4365 | 5 | 223840566 | 2,102395 |
| 794 | 1 | 513746829 | 0,17800706 | 794 | 1 | 513746829 | 0,1719258 | 4087 | 5 | 21770965 | 2,099218 |
| 640 | 1 | 410550498 | 0,17759370 | 640 | 1 | 410550498 | 0,1716950 | 933 | 1 | 606891568 | 2,090111 |
| 3964 | 4 | 247813127 | 0,17648493 | 3964 | 4 | 247813127 | 0,1710750 | 3044 | 4 | 35384247 | 2,088211 |
| 1402 | 2 | 249156009 | 0,17233340 | 1402 | 2 | 249156009 | 0,1687406 | 620 | 1 | 397148378 | 2,085016 |
| 3766 | 4 | 246990437 | 0,17221799 | 3766 | 4 | 246990437 | 0,1686755 | 4838 | 5 | 661898164 | 2,083066 |
| 966 | 1 | 629005068 | 0,17209854 | 966 | 1 | 629005068 | 0,1686080 | 5175 | 6 | 151958886 | 2,081309 |
| 2221 | 3 | 178234880 | 0,17189332 | 2221 | 3 | 178234880 | 0,1684920 | 2480 | 3 | 345824384 | 2,068159 |
| 1268 | 2 | 169989408 | 0,17037823 | 1268 | 2 | 169989408 | 0,1676343 | 3532 | 4 | 246018167 | 2,065721 |
| 4085 | 5 | 20200987 | 0,16673653 | 4085 | 5 | 20200987 | 0,1655611 | 5789 | 6 | 373022970 | 2,064492 |
| 629 | 1 | 403179332 | 0,16601164 | 629 | 1 | 403179332 | 0,1651465 | 2584 | 3 | 442569240 | 2,058015 |
| 5269 | 6 | 187827570 | 0,16512636 | 5269 | 6 | 187827570 | 0,1646392 | 3881 | 4 | 247468262 | 2,057619 |
| 2613 | 3 | 469546171 | 0,16461268 | 2613 | 3 | 469546171 | 0,1643445 | 5756 | 6 | 361270185 | 2,037824 |
| 1774 | 2 | 514982031 | 0,15771612 | 1774 | 2 | 514982031 | 0,1603536 | 1657 | 2 | 431375459 | 2,033325 |
| 5538 | 6 | 283630575 | 0,15729315 | 5538 | 6 | 283630575 | 0,1601067 | 4604 | 5 | 445184680 | 2,018235 |
| 414 | 1 | 267615649 | 0,15668913 | 414 | 1 | 267615649 | 0,1597538 | 2968 | 3 | 799781016 | 2,013910 |
| 37 | 1 | 9300541 | 0,15522200 | 37 | 1 | 9300541 | 0,1588946 | 2221 | 3 | 178234880 | 2,001622 |

Fonte: Da autora (2019)

Tabela 42 – Bayes B modificado \times GWAS
cenário Poligênico, $h^2 = 0,5$

| Bayes B modificado | | | | | | | | GWAS | | | |
|--------------------|----|-----------|-------------|-------|----|-----------|--------------------------------|-------|----|-----------|--------------------------------|
| Marca | Cr | Pos | wald | Marca | Cr | Pos | $-\log_{10}(\text{valor} - p)$ | Marca | Cr | Pos | $-\log_{10}(\text{valor} - p)$ |
| 5826 | 6 | 386200335 | 49,24104115 | 5826 | 6 | 386200335 | 11,6451911 | 5826 | 6 | 386200335 | 8,459814 |
| 3480 | 4 | 245802107 | 31,01340838 | 3480 | 4 | 245802107 | 7,5913327 | 3480 | 4 | 245802107 | 6,266500 |
| 1562 | 2 | 363489782 | 30,44027772 | 1562 | 2 | 363489782 | 7,4630572 | 1562 | 2 | 363489782 | 5,390297 |
| 3261 | 4 | 212893509 | 11,54759547 | 3261 | 4 | 212893509 | 3,1685338 | 3261 | 4 | 212893509 | 4,081755 |
| 4317 | 5 | 204516414 | 10,88203386 | 4317 | 5 | 204516414 | 3,0127738 | 5673 | 6 | 331710150 | 3,983664 |
| 2517 | 3 | 380243227 | 7,20009637 | 2517 | 3 | 380243227 | 2,1372745 | 3517 | 4 | 245955842 | 3,790299 |
| 1093 | 2 | 20954874 | 7,19851313 | 1093 | 2 | 20954874 | 2,1368913 | 5846 | 6 | 393323235 | 3,772569 |
| 3517 | 4 | 245955842 | 3,44613996 | 3517 | 4 | 245955842 | 1,1979130 | 2517 | 3 | 380243227 | 3,400196 |
| 4629 | 5 | 468337830 | 2,60520930 | 4629 | 5 | 468337830 | 0,9725970 | 5933 | 6 | 424307850 | 3,338913 |
| 514 | 1 | 293632633 | 1,71340220 | 514 | 1 | 293632633 | 0,7200035 | 5277 | 6 | 190676730 | 3,331761 |
| 4913 | 5 | 731357614 | 1,62348144 | 4913 | 5 | 731357614 | 0,6933455 | 1646 | 2 | 423515013 | 3,277031 |
| 1417 | 2 | 259874800 | 1,23850224 | 1417 | 2 | 259874800 | 0,5755105 | 3736 | 4 | 246865787 | 3,273612 |
| 4337 | 5 | 209498142 | 1,16421942 | 4337 | 5 | 209498142 | 0,5519244 | 5884 | 6 | 406856745 | 3,248133 |
| 4987 | 5 | 799890938 | 0,81956344 | 4987 | 5 | 799890938 | 0,4373410 | 4317 | 5 | 204516414 | 3,192595 |
| 2346 | 3 | 223513366 | 0,80303364 | 2346 | 3 | 223513366 | 0,4315778 | 4649 | 5 | 486860350 | 3,177421 |
| 187 | 1 | 83595286 | 0,73702665 | 187 | 1 | 83595286 | 0,4082511 | 2749 | 3 | 596058675 | 3,177365 |
| 5577 | 6 | 297520230 | 0,69820499 | 5577 | 6 | 297520230 | 0,3942776 | 5862 | 6 | 399021555 | 3,117959 |
| 4549 | 5 | 394247750 | 0,64584901 | 4549 | 5 | 394247750 | 0,3750987 | 3381 | 4 | 245390762 | 3,035844 |
| 4341 | 5 | 210890745 | 0,62211442 | 4341 | 5 | 210890745 | 0,3662664 | 5670 | 6 | 330641715 | 3,008389 |
| 3773 | 4 | 247019522 | 0,61781035 | 3773 | 4 | 247019522 | 0,3646549 | 187 | 1 | 83595286 | 2,927185 |
| 5673 | 6 | 331710150 | 0,58479188 | 5673 | 6 | 331710150 | 0,3521867 | 5799 | 6 | 376584420 | 2,906733 |
| 3971 | 4 | 247842212 | 0,52277116 | 3971 | 4 | 247842212 | 0,3282146 | 1093 | 2 | 20954874 | 2,879809 |
| 1534 | 2 | 343481372 | 0,51944231 | 1534 | 2 | 343481372 | 0,3269056 | 1865 | 2 | 580009364 | 2,858210 |
| 5846 | 6 | 393323235 | 0,45485969 | 5846 | 6 | 393323235 | 0,3009986 | 3389 | 4 | 245424002 | 2,708215 |
| 5277 | 6 | 190676730 | 0,42869074 | 5277 | 6 | 190676730 | 0,2901942 | 5033 | 6 | 75187712 | 2,700129 |
| 4540 | 5 | 385912616 | 0,41618643 | 4540 | 5 | 385912616 | 0,2849614 | 5805 | 6 | 378721290 | 2,686021 |
| 1263 | 2 | 169265278 | 0,40189113 | 1263 | 2 | 169265278 | 0,2789200 | 427 | 1 | 268373293 | 2,590842 |
| 1295 | 2 | 186512163 | 0,37983718 | 1295 | 2 | 186512163 | 0,2694677 | 4290 | 5 | 193521948 | 2,568202 |
| 2088 | 3 | 69892547 | 0,37085962 | 2088 | 3 | 69892547 | 0,2655714 | 945 | 1 | 614932841 | 2,566545 |
| 1230 | 2 | 144528591 | 0,36273909 | 1230 | 2 | 144528591 | 0,2620215 | 5698 | 6 | 340613775 | 2,563206 |
| 3960 | 4 | 247796507 | 0,35262476 | 3960 | 4 | 247796507 | 0,2575647 | 4265 | 5 | 184588958 | 2,491110 |
| 2800 | 3 | 643500864 | 0,31969859 | 2800 | 3 | 643500864 | 0,2427643 | 4838 | 5 | 661898164 | 2,438240 |
| 5461 | 6 | 256207410 | 0,31961487 | 5461 | 6 | 256207410 | 0,2427261 | 4453 | 5 | 305396584 | 2,422658 |
| 5862 | 6 | 399021555 | 0,30241314 | 5862 | 6 | 399021555 | 0,2347984 | 1095 | 2 | 20958307 | 2,421069 |
| 1646 | 2 | 423515013 | 0,29970249 | 1646 | 2 | 423515013 | 0,2335358 | 3133 | 4 | 128632208 | 2,401665 |
| 5962 | 6 | 434636055 | 0,29464849 | 5962 | 6 | 434636055 | 0,2311718 | 3096 | 4 | 76123039 | 2,394804 |
| 4081 | 5 | 19973330 | 0,29353605 | 4081 | 5 | 19973330 | 0,2306497 | 5619 | 6 | 312478320 | 2,369837 |
| 4138 | 5 | 62622237 | 0,29056509 | 4138 | 5 | 62622237 | 0,2292521 | 2877 | 3 | 715129267 | 2,355939 |
| 1095 | 2 | 20958307 | 0,27858235 | 1095 | 2 | 20958307 | 0,2235662 | 3971 | 4 | 247842212 | 2,332747 |
| 1321 | 2 | 198757428 | 0,27781504 | 1321 | 2 | 198757428 | 0,2231994 | 4805 | 5 | 628557628 | 2,332568 |
| 5670 | 6 | 330641715 | 0,27430419 | 5670 | 6 | 330641715 | 0,2215166 | 4648 | 5 | 485934224 | 2,326017 |
| 4266 | 5 | 185596985 | 0,27240698 | 4266 | 5 | 185596985 | 0,2206043 | 5244 | 6 | 178923945 | 2,324638 |
| 5884 | 6 | 406856745 | 0,27005069 | 5884 | 6 | 406856745 | 0,2194683 | 1358 | 2 | 214423504 | 2,304391 |
| 4098 | 5 | 30239292 | 0,27000393 | 4098 | 5 | 30239292 | 0,2194457 | 3972 | 4 | 247846367 | 2,286949 |
| 671 | 1 | 431323786 | 0,25351717 | 671 | 1 | 431323786 | 0,2114008 | 499 | 1 | 291031959 | 2,284529 |
| 1561 | 2 | 362775196 | 0,24882889 | 1561 | 2 | 362775196 | 0,2090812 | 1185 | 2 | 58354695 | 2,283718 |
| 4648 | 5 | 485934224 | 0,24678274 | 4648 | 5 | 485934224 | 0,2080641 | 5103 | 6 | 113384656 | 2,265262 |
| 2946 | 3 | 779315758 | 0,24656584 | 2946 | 3 | 779315758 | 0,2079561 | 1321 | 2 | 198757428 | 2,264077 |
| 970 | 1 | 631685492 | 0,24181254 | 970 | 1 | 631685492 | 0,2055816 | 821 | 1 | 531839692 | 2,263088 |
| 100 | 1 | 34479054 | 0,22101707 | 100 | 1 | 34479054 | 0,1949983 | 308 | 1 | 216989434 | 2,252107 |
| 4838 | 5 | 661898164 | 0,21231803 | 4838 | 5 | 661898164 | 0,1904695 | 1354 | 2 | 214263400 | 2,231001 |
| 1167 | 2 | 44606542 | 0,20934392 | 1167 | 2 | 44606542 | 0,1889064 | 5547 | 6 | 286835880 | 2,222143 |
| 750 | 1 | 484262163 | 0,20690024 | 750 | 1 | 484262163 | 0,1876162 | 4629 | 5 | 468337830 | 2,186507 |
| 43 | 1 | 10067746 | 0,20236364 | 43 | 1 | 10067746 | 0,1852067 | 4170 | 5 | 75234294 | 2,178396 |
| 4170 | 5 | 75234294 | 0,20125349 | 4170 | 5 | 75234294 | 0,1846142 | 3992 | 4 | 247929467 | 2,174125 |
| 5050 | 6 | 80877074 | 0,20121979 | 5050 | 6 | 80877074 | 0,1845962 | 3005 | 4 | 2637306 | 2,168391 |
| 4670 | 5 | 506308996 | 0,19787911 | 4670 | 5 | 506308996 | 0,1828061 | 2072 | 3 | 53549251 | 2,162393 |
| 945 | 1 | 614932841 | 0,19261163 | 945 | 1 | 614932841 | 0,1799616 | 1417 | 2 | 259874800 | 2,154835 |
| 1354 | 2 | 214263400 | 0,18496200 | 1354 | 2 | 214263400 | 0,1757809 | 5789 | 6 | 373022970 | 2,123864 |
| 4205 | 5 | 145121333 | 0,18423926 | 4205 | 5 | 145121333 | 0,1753828 | 1561 | 2 | 362775196 | 2,111705 |

Fonte: Da autora (2019)

Tabela 43 – Bayes B modificado \times GWAS
cenário Poligênico, $h^2 = 0,8$

| Bayes B modificado | | | | | | | | GWAS | | | |
|--------------------|----|-----------|------------|-------|----|-----------|--------------------------------|-------|----|-----------|--------------------------------|
| Marca | Cr | Pos | wald | Marca | Cr | Pos | $-\log_{10}(\text{valor} - p)$ | Marca | Cr | Pos | $-\log_{10}(\text{valor} - p)$ |
| 5826 | 6 | 386200335 | 95,9968557 | 5826 | 6 | 386200335 | 21,9390527 | 5826 | 6 | 386200335 | 11,587916 |
| 3480 | 4 | 245802107 | 60,6122248 | 3480 | 4 | 245802107 | 14,1580086 | 3480 | 4 | 245802107 | 8,318912 |
| 1562 | 2 | 363489782 | 56,2670242 | 1562 | 2 | 363489782 | 13,1988183 | 5954 | 6 | 431786895 | 5,722483 |
| 2517 | 3 | 380243227 | 25,5207743 | 2517 | 3 | 380243227 | 6,3588778 | 1562 | 2 | 363489782 | 5,662491 |
| 3261 | 4 | 212893509 | 25,2759594 | 3261 | 4 | 212893509 | 6,3037632 | 5033 | 6 | 75187712 | 5,350154 |
| 4549 | 5 | 394247750 | 21,6306717 | 4549 | 5 | 394247750 | 5,4807972 | 5862 | 6 | 399021555 | 4,623745 |
| 1093 | 2 | 20954874 | 19,9307943 | 1093 | 2 | 20954874 | 5,0953033 | 2517 | 3 | 380243227 | 4,065757 |
| 4629 | 5 | 468337830 | 19,8147830 | 4629 | 5 | 468337830 | 5,0689479 | 5590 | 6 | 302150115 | 3,514904 |
| 3960 | 4 | 247796507 | 16,9790909 | 3960 | 4 | 247796507 | 4,4225801 | 3736 | 4 | 246865787 | 3,498160 |
| 3288 | 4 | 235779015 | 8,8439096 | 3288 | 4 | 235779015 | 2,5315515 | 1093 | 2 | 20954874 | 3,365941 |
| 5560 | 6 | 291465765 | 3,5772974 | 5560 | 6 | 291465765 | 1,2322924 | 2292 | 3 | 200201258 | 3,072857 |
| 2289 | 3 | 200142461 | 3,3307091 | 2289 | 3 | 200142461 | 1,1675067 | 3517 | 4 | 245955842 | 3,015800 |
| 2279 | 3 | 194643731 | 3,3026165 | 2279 | 3 | 194643731 | 1,1600846 | 859 | 1 | 557303721 | 2,967605 |
| 5862 | 6 | 399021555 | 2,8934464 | 5862 | 6 | 399021555 | 1,0509005 | 3643 | 4 | 246479372 | 2,908547 |
| 5033 | 6 | 75187712 | 2,7504448 | 5033 | 6 | 75187712 | 1,0122114 | 3996 | 4 | 247946087 | 2,879696 |
| 3476 | 4 | 245785487 | 2,6839788 | 3476 | 4 | 245785487 | 0,9941240 | 89 | 1 | 28084996 | 2,879099 |
| 3337 | 4 | 245207942 | 1,8786780 | 3337 | 4 | 245207942 | 0,7683152 | 3918 | 4 | 247621997 | 2,809284 |
| 5643 | 6 | 321025800 | 1,7804538 | 5643 | 6 | 321025800 | 0,7397053 | 5134 | 6 | 131576099 | 2,790126 |
| 3318 | 4 | 245131700 | 1,6231799 | 3318 | 4 | 245131700 | 0,6932556 | 1865 | 2 | 580009364 | 2,786465 |
| 4648 | 5 | 485934224 | 1,5682930 | 4648 | 5 | 485934224 | 0,6768406 | 4629 | 5 | 468337830 | 2,781031 |
| 1528 | 2 | 339193855 | 1,2952800 | 1528 | 2 | 339193855 | 0,5333284 | 5636 | 6 | 318532785 | 2,764711 |
| 5109 | 6 | 113385418 | 1,1914368 | 5109 | 6 | 113385418 | 0,5606043 | 1304 | 2 | 189293647 | 2,700383 |
| 3252 | 4 | 206492925 | 1,1220993 | 3252 | 4 | 206492925 | 0,5384011 | 5186 | 6 | 157129197 | 2,698615 |
| 4085 | 5 | 20200987 | 1,0477805 | 4085 | 5 | 20200987 | 0,5142519 | 5063 | 6 | 83064740 | 2,698013 |
| 5590 | 6 | 302150115 | 1,0319118 | 5590 | 6 | 302150115 | 0,5090447 | 5805 | 6 | 378721290 | 2,687730 |
| 2736 | 3 | 583965568 | 1,0141447 | 2736 | 3 | 583965568 | 0,5031922 | 3476 | 4 | 245785487 | 2,630147 |
| 5472 | 6 | 260125005 | 0,9703652 | 5472 | 6 | 260125005 | 0,4886660 | 1646 | 2 | 423515013 | 2,593889 |
| 1225 | 2 | 140264872 | 0,8796760 | 1225 | 2 | 140264872 | 0,4580585 | 5801 | 6 | 377296710 | 2,582850 |
| 382 | 1 | 262715518 | 0,8097848 | 382 | 1 | 262715518 | 0,4339352 | 2840 | 3 | 680710424 | 2,569747 |
| 5636 | 6 | 318532785 | 0,7293474 | 5636 | 6 | 318532785 | 0,4055027 | 966 | 1 | 629005068 | 2,547537 |
| 1087 | 2 | 20042233 | 0,6626158 | 1087 | 2 | 20042233 | 0,3812848 | 5530 | 6 | 280781415 | 2,532024 |
| 3488 | 4 | 245835347 | 0,6286720 | 3488 | 4 | 245835347 | 0,3687157 | 4459 | 5 | 310896410 | 2,531509 |
| 3406 | 4 | 245494637 | 0,6240501 | 3406 | 4 | 245494637 | 0,3669901 | 5416 | 6 | 240180885 | 2,488117 |
| 3755 | 4 | 246944732 | 0,5919590 | 3755 | 4 | 246944732 | 0,3549094 | 1933 | 2 | 628601218 | 2,436105 |
| 4951 | 5 | 766550402 | 0,5888471 | 4951 | 5 | 766550402 | 0,3537284 | 3658 | 4 | 246541697 | 2,433635 |
| 4459 | 5 | 310896410 | 0,5831946 | 4459 | 5 | 310896410 | 0,3515786 | 3538 | 4 | 246043097 | 2,42886 |
| 4849 | 5 | 672085550 | 0,5791369 | 4849 | 5 | 672085550 | 0,3500318 | 5442 | 6 | 249440655 | 2,404876 |
| 89 | 1 | 28084996 | 0,5731609 | 89 | 1 | 28084996 | 0,3477483 | 3220 | 4 | 180309373 | 2,391637 |
| 4895 | 5 | 714687346 | 0,5392302 | 4895 | 5 | 714687346 | 0,3346518 | 5290 | 6 | 195306615 | 2,384984 |
| 1646 | 2 | 423515013 | 0,5283911 | 1646 | 2 | 423515013 | 0,3304190 | 382 | 1 | 262715518 | 2,379969 |
| 251 | 1 | 183647544 | 0,5259151 | 251 | 1 | 183647544 | 0,3294487 | 5782 | 6 | 370529955 | 2,373650 |
| 3350 | 4 | 245261957 | 0,5082205 | 3350 | 4 | 245261957 | 0,3224750 | 1086 | 2 | 19837545 | 2,371760 |
| 4820 | 5 | 645227896 | 0,5081494 | 4820 | 5 | 645227896 | 0,3224468 | 4549 | 5 | 394247750 | 2,352150 |
| 3679 | 4 | 246628952 | 0,4828097 | 3679 | 4 | 246628952 | 0,3123352 | 2998 | 3 | 827688186 | 2,352145 |
| 5417 | 6 | 240537030 | 0,4581732 | 5417 | 6 | 240537030 | 0,3023531 | 1880 | 2 | 590728155 | 2,342969 |
| 3205 | 4 | 173687725 | 0,4547807 | 3205 | 4 | 173687725 | 0,3009663 | 5698 | 6 | 340613775 | 2,338420 |
| 3845 | 4 | 247318682 | 0,4469014 | 3845 | 4 | 247318682 | 0,2977331 | 3295 | 4 | 238553534 | 2,331521 |
| 3682 | 4 | 246641417 | 0,4249733 | 3682 | 4 | 246641417 | 0,2886435 | 5171 | 6 | 149251173 | 2,318388 |
| 4266 | 5 | 185596985 | 0,4230924 | 4266 | 5 | 185596985 | 0,2878573 | 3337 | 4 | 245207942 | 2,312483 |
| 522 | 1 | 294646235 | 0,4194458 | 522 | 1 | 294646235 | 0,2863300 | 728 | 1 | 469519830 | 2,297564 |
| 5518 | 6 | 276507675 | 0,4102214 | 5518 | 6 | 276507675 | 0,2824484 | 1686 | 2 | 452098456 | 2,294555 |
| 5461 | 6 | 256207410 | 0,3826016 | 5461 | 6 | 256207410 | 0,2706617 | 3632 | 4 | 246433667 | 2,263630 |
| 1249 | 2 | 161879250 | 0,3505345 | 1249 | 2 | 161879250 | 0,2566386 | 3261 | 4 | 212893509 | 2,223022 |
| 3898 | 4 | 247538897 | 0,3434822 | 3898 | 4 | 247538897 | 0,2535012 | 5641 | 6 | 320313510 | 2,207166 |
| 5416 | 6 | 240180885 | 0,3377537 | 5416 | 6 | 240180885 | 0,2509375 | 5244 | 6 | 178923945 | 2,205616 |
| 886 | 1 | 575396584 | 0,3306048 | 886 | 1 | 575396584 | 0,2477186 | 4849 | 5 | 672085550 | 2,204546 |
| 4853 | 5 | 675790054 | 0,3277923 | 4853 | 5 | 675790054 | 0,2464461 | 1249 | 2 | 161879250 | 2,186632 |
| 1153 | 2 | 39041117 | 0,3257627 | 1153 | 2 | 39041117 | 0,2455256 | 5336 | 6 | 211689285 | 2,170129 |
| 3032 | 4 | 28192994 | 0,3256827 | 3032 | 4 | 28192994 | 0,2454893 | 4394 | 5 | 250698220 | 2,163688 |
| 233 | 1 | 173860693 | 0,3174186 | 233 | 1 | 173860693 | 0,2417218 | 1867 | 2 | 581438536 | 2,150772 |

Fonte: Da autora (2019)

Tabelas BSSV modificado x GWAS

Tabela 44 – BSSV modificado × GWAS
cenário Oligogênico, $h^2 = 0,2$

| BSSV modificado | | | | GWAS | | | |
|-----------------|----|---|-------------|-------|----|--------------------------------|--------------------------------|
| wald | | $-\log_{10}(\text{valor} - p)$ teste Wald | | wald | | $-\log_{10}(\text{valor} - p)$ | |
| Marca | Cr | Pos | wald | Marca | Cr | Pos | $-\log_{10}(\text{valor} - p)$ |
| 4004 | 5 | 2123914 | 31,30926866 | 4004 | 5 | 2123914 | 7,65752463 |
| 2983 | 3 | 813734601 | 9,07547504 | 2983 | 3 | 813734601 | 2,58659919 |
| 2897 | 3 | 733734047 | 3,54231046 | 2897 | 3 | 733734047 | 1,22313851 |
| 1029 | 2 | 4679573 | 1,97710048 | 1029 | 2 | 4679573 | 0,79670542 |
| 2896 | 3 | 732803808 | 1,05837903 | 2896 | 3 | 732803808 | 0,51771943 |
| 279 | 1 | 202620610 | 1,00250595 | 279 | 1 | 202620610 | 0,49934522 |
| 4877 | 5 | 698017078 | 0,78419186 | 4877 | 5 | 698017078 | 0,42497137 |
| 4045 | 5 | 9279776 | 0,70185877 | 4045 | 5 | 9279776 | 0,39560132 |
| 922 | 1 | 599520402 | 0,68471463 | 922 | 1 | 599520402 | 0,38937393 |
| 4889 | 5 | 709130590 | 0,60771323 | 4889 | 5 | 709130590 | 0,36086216 |
| 1826 | 2 | 552140507 | 0,56847077 | 1826 | 2 | 552140507 | 0,34595133 |
| 4004 | 5 | 2123914 | | 4004 | 5 | 2123914 | 7,597231 |
| 2983 | 3 | 813734601 | | 2983 | 3 | 813734601 | 4,887828 |
| 2897 | 3 | 733734047 | | 2897 | 3 | 733734047 | 4,265723 |
| 1029 | 2 | 4679573 | | 1029 | 2 | 4679573 | 3,633500 |
| 2896 | 3 | 732803808 | | 2896 | 3 | 732803808 | 3,457414 |
| 279 | 1 | 202620610 | | 4303 | 5 | 200813183 | 3,316936 |
| 4877 | 5 | 698017078 | | 4877 | 5 | 698017078 | 3,296434 |
| 4045 | 5 | 9279776 | | 2897 | 3 | 733734047 | 3,229439 |
| 922 | 1 | 599520402 | | 4047 | 5 | 9965534 | 3,221998 |
| 4889 | 5 | 709130590 | | 4621 | 5 | 460928822 | 3,214396 |
| 1826 | 2 | 552140507 | | 279 | 1 | 202620610 | 3,198296 |

Fonte: Da autora (2019)

Tabela 45 – BSSV modificado × GWAS
cenário Oligogênico, $h^2 = 0,5$

| BSSV modificado | | | | GWAS | | | |
|-----------------|----|---|-------------|-------|----|--------------------------------|--------------------------------|
| wald | | $-\log_{10}(\text{valor} - p)$ teste Wald | | wald | | $-\log_{10}(\text{valor} - p)$ | |
| Marca | Cr | Pos | wald | Marca | Cr | Pos | $-\log_{10}(\text{valor} - p)$ |
| 4004 | 5 | 2123914 | 80,75659533 | 4004 | 5 | 2123914 | 18,59294150 |
| 3677 | 4 | 246620642 | 67,25973133 | 3677 | 4 | 246620642 | 15,62343551 |
| 4877 | 5 | 698017078 | 56,87172385 | 4877 | 5 | 698017078 | 13,33237271 |
| 193 | 1 | 88380847 | 6,97646404 | 193 | 1 | 88380847 | 2,08307974 |
| 1968 | 2 | 653611730 | 1,78489783 | 1968 | 2 | 653611730 | 0,74100602 |
| 877 | 1 | 569365630 | 0,90817578 | 877 | 1 | 569365630 | 0,46775676 |
| 713 | 1 | 459468239 | 0,87727681 | 713 | 1 | 459468239 | 0,45723851 |
| 532 | 1 | 297315711 | 0,85296524 | 532 | 1 | 297315711 | 0,44889802 |
| 538 | 1 | 298413091 | 0,80539649 | 538 | 1 | 298413091 | 0,43240343 |
| 833 | 1 | 539880964 | 0,63722285 | 833 | 1 | 539880964 | 0,37189909 |
| 4004 | 5 | 2123914 | | 4004 | 5 | 2123914 | 13,339721 |
| 3677 | 4 | 246620642 | | 3677 | 4 | 246620642 | 9,102989 |
| 4877 | 5 | 698017078 | | 4877 | 5 | 698017078 | 7,418931 |
| 193 | 1 | 88380847 | | 3805 | 4 | 247152482 | 5,668896 |
| 1968 | 2 | 653611730 | | 4260 | 5 | 181672656 | 5,507352 |
| 877 | 1 | 569365630 | | 4625 | 5 | 464633326 | 5,215285 |
| 713 | 1 | 459468239 | | 3933 | 4 | 247684322 | 4,264209 |
| 532 | 1 | 297315711 | | 3549 | 4 | 246088802 | 3,997775 |
| 538 | 1 | 298413091 | | 4772 | 5 | 600773848 | 3,850929 |
| 833 | 1 | 539880964 | | 3037 | 4 | 29209539 | 3,656472 |

Fonte: Da autora (2019)

Tabela 46 – BSSV modificado × GWAS
cenário Oligogênico, $h^2 = 0,8$

| BSSV modificado | | | | GWAS | | | |
|-----------------|----|---|--------------|-------|----|--------------------------------|--------------------------------|
| wald | | $-\log_{10}(\text{valor} - p)$ teste Wald | | wald | | $-\log_{10}(\text{valor} - p)$ | |
| Marca | Cr | Pos | wald | Marca | Cr | Pos | $-\log_{10}(\text{valor} - p)$ |
| 4004 | 5 | 2123914 | 379,79438517 | 4004 | 5 | 2123914 | 83,86027314 |
| 4877 | 5 | 698017078 | 187,54292186 | 4877 | 5 | 698017078 | 41,96132394 |
| 3677 | 4 | 246620642 | 96,84182130 | 3677 | 4 | 246620642 | 22,12440002 |
| 538 | 1 | 298413091 | 25,80681706 | 538 | 1 | 298413091 | 6,42325182 |
| 2088 | 3 | 69892547 | 23,50107453 | 2088 | 3 | 69892547 | 5,90363225 |
| 2000 | 2 | 676478485 | 6,98242431 | 2000 | 2 | 676478485 | 2,08452614 |
| 2855 | 3 | 694664009 | 0,67999135 | 2855 | 3 | 694664009 | 0,38765097 |
| 4927 | 5 | 744323378 | 0,62240861 | 4927 | 5 | 744323378 | 0,36637640 |
| 5764 | 6 | 364119345 | 0,59905878 | 5764 | 6 | 364119345 | 0,35759744 |
| 900 | 1 | 584778069 | 0,49086528 | 900 | 1 | 584778069 | 0,31556618 |
| 4004 | 5 | 2123914 | | 4004 | 5 | 2123914 | 38,278105 |
| 4877 | 5 | 698017078 | | 4877 | 5 | 698017078 | 15,782135 |
| 3677 | 4 | 246620642 | | 4260 | 5 | 181672656 | 10,589633 |
| 538 | 1 | 298413091 | | 3677 | 4 | 246620642 | 6,920932 |
| 2088 | 3 | 69892547 | | 4303 | 5 | 200813183 | 6,386065 |
| 2000 | 2 | 676478485 | | 4237 | 5 | 171688991 | 5,372977 |
| 2855 | 3 | 694664009 | | 5412 | 6 | 238756305 | 4,982878 |
| 4927 | 5 | 744323378 | | 4625 | 5 | 464633326 | 4,885838 |
| 5764 | 6 | 364119345 | | 4047 | 5 | 9965534 | 4,743689 |
| 900 | 1 | 584778069 | | 4772 | 5 | 600773848 | 3,910513 |

Fonte: Da autora (2019)

Tabela 47 – BSSV modificado \times GWAS
cenário Poligênico, $h^2 = 0,2$

| BSSV modificado | | | | | | | GWAS | | | | |
|-----------------|----|-----------|---|-------|----|-----------|--------------------------------|-------|----|-----------|--------------------------------|
| wald | | | $-\log_{10}(\text{valor} - p)$ teste Wald | | | | $-\log_{10}(\text{valor} - p)$ | | | | |
| Marca | Cr | Pos | wald | Marca | Cr | Pos | $-\log_{10}(\text{valor} - p)$ | Marca | Cr | Pos | $-\log_{10}(\text{valor} - p)$ |
| 5826 | 6 | 386200335 | 5,75958301 | 5826 | 6 | 386200335 | 1,7851836 | 5826 | 6 | 386200335 | 4,358662 |
| 3480 | 4 | 245802107 | 4,89138431 | 3480 | 4 | 245802107 | 1,5687804 | 3480 | 4 | 245802107 | 4,130539 |
| 5862 | 6 | 399021555 | 2,46366767 | 5862 | 6 | 399021555 | 0,9336500 | 3941 | 4 | 247717562 | 3,956526 |
| 3517 | 4 | 245955842 | 1,78012636 | 3517 | 4 | 245955842 | 0,7396094 | 5862 | 6 | 399021555 | 3,582077 |
| 5728 | 6 | 351298125 | 1,59811352 | 5728 | 6 | 351298125 | 0,6857729 | 4966 | 5 | 780442292 | 3,359942 |
| 2840 | 3 | 680710424 | 1,50119528 | 2840 | 3 | 680710424 | 0,6566159 | 5801 | 6 | 377296710 | 3,308852 |
| 1408 | 2 | 253443525 | 1,25659064 | 1408 | 2 | 253443525 | 0,5812060 | 2840 | 3 | 680710424 | 3,179535 |
| 5801 | 6 | 377296710 | 1,25445347 | 5801 | 6 | 377296710 | 0,5805340 | 5239 | 6 | 177143220 | 3,076366 |
| 3941 | 4 | 247717562 | 1,19852000 | 3941 | 4 | 247717562 | 0,4310551 | 3517 | 4 | 245955842 | 3,035031 |
| 4966 | 5 | 780442292 | 1,07316690 | 4966 | 5 | 780442292 | 0,5225442 | 1408 | 2 | 253443525 | 2,992735 |
| 1562 | 2 | 363489782 | 1,01906919 | 1562 | 2 | 363489782 | 0,5048167 | 5123 | 6 | 121839487 | 2,850738 |
| 5239 | 6 | 177143220 | 0,80847026 | 5239 | 6 | 177143220 | 0,4334766 | 3658 | 4 | 246541697 | 2,838595 |
| 5269 | 6 | 187827570 | 0,80153891 | 5269 | 6 | 187827570 | 0,4310551 | 3810 | 4 | 247173257 | 2,817565 |
| 3658 | 4 | 246541697 | 0,76823086 | 3658 | 4 | 246541697 | 0,4193431 | 805 | 1 | 521117995 | 2,805084 |
| 3718 | 4 | 246790997 | 0,70721872 | 3718 | 4 | 246790997 | 0,3975399 | 3515 | 4 | 245947532 | 2,764010 |
| 620 | 1 | 397148378 | 0,63261105 | 620 | 1 | 397148378 | 0,3701836 | 3766 | 4 | 246990437 | 2,738281 |
| 805 | 1 | 521117995 | 0,58120690 | 805 | 1 | 521117995 | 0,3508213 | 3420 | 4 | 245552807 | 2,708764 |
| 3766 | 4 | 246990437 | 0,57787229 | 3766 | 4 | 246990437 | 0,3495492 | 5728 | 6 | 351298125 | 2,683066 |
| 5123 | 6 | 121839487 | 0,56909387 | 5123 | 6 | 121839487 | 0,3461903 | 5269 | 6 | 187827570 | 2,602368 |
| 2578 | 3 | 436987806 | 0,47666581 | 2578 | 3 | 436987806 | 0,3098602 | 1562 | 2 | 363489782 | 2,588188 |
| 842 | 1 | 545911919 | 0,46750056 | 842 | 1 | 545911919 | 0,3061506 | 2578 | 3 | 436987806 | 2,492446 |
| 4058 | 5 | 11737498 | 0,45190830 | 4058 | 5 | 11737498 | 0,2997896 | 1573 | 2 | 371350229 | 2,436171 |
| 1268 | 2 | 169989408 | 0,43438052 | 1268 | 2 | 169989408 | 0,2925599 | 3718 | 4 | 246790997 | 2,397791 |
| 1573 | 2 | 371350229 | 0,41883739 | 1573 | 2 | 371350229 | 0,2860748 | 3960 | 4 | 247796507 | 2,396755 |
| 3296 | 4 | 239237605 | 0,39381778 | 3296 | 4 | 239237605 | 0,2754789 | 2025 | 3 | 9124190 | 2,390213 |
| 3810 | 4 | 247173257 | 0,39355081 | 3810 | 4 | 247173257 | 0,2753647 | 1941 | 2 | 634317906 | 2,383527 |
| 3515 | 4 | 245947532 | 0,37881939 | 3515 | 4 | 245947532 | 0,2690275 | 4582 | 5 | 424809908 | 2,381976 |
| 4087 | 5 | 21770965 | 0,35039515 | 4087 | 5 | 21770965 | 0,2565768 | 1268 | 2 | 169989408 | 2,359644 |
| 1657 | 2 | 431375459 | 0,34469553 | 1657 | 2 | 431375459 | 0,2540424 | 5763 | 6 | 363763200 | 2,356424 |
| 5210 | 6 | 165724924 | 0,31307736 | 5210 | 6 | 165724924 | 0,2397299 | 5933 | 6 | 424307850 | 2,329228 |
| 1941 | 2 | 634317906 | 0,30506510 | 1941 | 2 | 634317906 | 0,2360300 | 4070 | 5 | 13981667 | 2,307127 |
| 3960 | 4 | 247796507 | 0,29127860 | 3960 | 4 | 247796507 | 0,2295881 | 5990 | 6 | 444608115 | 2,297095 |
| 2726 | 3 | 574663178 | 0,25461856 | 2726 | 3 | 574663178 | 0,2119437 | 842 | 1 | 545911919 | 2,284275 |
| 1406 | 2 | 252014353 | 0,25419943 | 1406 | 2 | 252014353 | 0,2117372 | 5954 | 6 | 431786895 | 2,269155 |
| 5990 | 6 | 444608115 | 0,24972986 | 5990 | 6 | 444608115 | 0,2095281 | 382 | 1 | 262715518 | 2,253707 |
| 3399 | 4 | 245465552 | 0,24277327 | 3399 | 4 | 245465552 | 0,2060628 | 5478 | 6 | 262261875 | 2,213185 |
| 3044 | 4 | 35384247 | 0,24080123 | 3044 | 4 | 35384247 | 0,2050744 | 5350 | 6 | 216675315 | 2,187097 |
| 382 | 1 | 262715518 | 0,22577596 | 382 | 1 | 262715518 | 0,1974495 | 2726 | 3 | 574663178 | 2,168420 |
| 3387 | 4 | 245415692 | 0,22312847 | 3387 | 4 | 245415692 | 0,1960880 | 5693 | 6 | 338833050 | 2,146788 |
| 4704 | 5 | 537797280 | 0,21454362 | 4704 | 5 | 537797280 | 0,1916342 | 3296 | 4 | 239237605 | 2,138550 |
| 4380 | 5 | 237732456 | 0,21435815 | 4380 | 5 | 237732456 | 0,1915373 | 4960 | 5 | 774885536 | 2,124039 |
| 4070 | 5 | 13981667 | 0,19976274 | 4070 | 5 | 13981667 | 0,1838168 | 3399 | 4 | 245465552 | 2,105284 |
| 2584 | 3 | 442569240 | 0,19790455 | 2584 | 3 | 442569240 | 0,1828198 | 2901 | 3 | 737455003 | 2,105209 |
| 3846 | 4 | 247322837 | 0,19787471 | 3846 | 4 | 247322837 | 0,1828038 | 4365 | 5 | 223840566 | 2,102395 |
| 3420 | 4 | 245552807 | 0,19712200 | 3420 | 4 | 245552807 | 0,1823989 | 4087 | 5 | 21770965 | 2,099218 |
| 5756 | 6 | 361270185 | 0,19668667 | 5756 | 6 | 361270185 | 0,1821646 | 933 | 1 | 606891568 | 2,090111 |
| 2613 | 3 | 469546171 | 0,19229995 | 2613 | 3 | 469546171 | 0,1797925 | 3044 | 4 | 35384247 | 2,088211 |
| 5478 | 6 | 262261875 | 0,18929628 | 5478 | 6 | 262261875 | 0,1781571 | 620 | 1 | 397148378 | 2,085016 |
| 3001 | 4 | 139753 | 0,18825306 | 3001 | 4 | 139753 | 0,1775870 | 4838 | 5 | 661898164 | 2,083066 |
| 1865 | 2 | 580009364 | 0,18656141 | 1865 | 2 | 580009364 | 0,1766601 | 5175 | 6 | 151958886 | 2,081309 |
| 2117 | 3 | 122632201 | 0,18483679 | 2117 | 3 | 122632201 | 0,1757120 | 2480 | 3 | 345824384 | 2,068159 |
| 629 | 1 | 403179332 | 0,18454591 | 629 | 1 | 403179332 | 0,1755518 | 3532 | 4 | 246018167 | 2,065721 |
| 3780 | 4 | 247048607 | 0,17724581 | 3780 | 4 | 247048607 | 0,1715006 | 5789 | 6 | 373022970 | 2,064492 |
| 37 | 1 | 9300541 | 0,17546624 | 37 | 1 | 9300541 | 0,1705040 | 2584 | 3 | 442569240 | 2,058015 |
| 5693 | 6 | 338833050 | 0,17541637 | 5693 | 6 | 338833050 | 0,1704761 | 3881 | 4 | 247468262 | 2,057619 |
| 5350 | 6 | 216675315 | 0,17446660 | 5350 | 6 | 216675315 | 0,1699426 | 5756 | 6 | 361270185 | 2,037824 |
| 3881 | 4 | 247468262 | 0,17313321 | 3881 | 4 | 247468262 | 0,1691919 | 1657 | 2 | 431375459 | 2,033325 |
| 1518 | 2 | 332047994 | 0,17231986 | 1518 | 2 | 332047994 | 0,1687330 | 4604 | 5 | 445184860 | 2,018235 |
| 4582 | 5 | 424809908 | 0,17213109 | 4582 | 5 | 424809908 | 0,1686264 | 2968 | 3 | 799781016 | 2,013910 |
| 3626 | 4 | 246408737 | 0,17168711 | 3626 | 4 | 246408737 | 0,1683754 | 2221 | 3 | 178234880 | 2,001622 |

Fonte: Da autora (2019)

Tabela 48 – BSSV modificado × GWAS
cenário Poligênico, $h^2 = 0,5$

| BSSV modificado | | | | | | | | GWAS | | | |
|-----------------|----|-----------|-------------|---|----|-----------|--------------------------------|--------------------------------|----|-----------|--------------------------------|
| wald | | | | $-\log_{10}(\text{valor} - p)$ teste Wald | | | | $-\log_{10}(\text{valor} - p)$ | | | |
| Marca | Cr | Pos | wald | Marca | Cr | Pos | $-\log_{10}(\text{valor} - p)$ | Marca | Cr | Pos | $-\log_{10}(\text{valor} - p)$ |
| 5826 | 6 | 386200335 | 46,53570979 | 5826 | 6 | 386200335 | 11,0459315 | 5826 | 6 | 386200335 | 8,459814 |
| 3480 | 4 | 245802107 | 28,90558140 | 3480 | 4 | 245802107 | 7,1192229 | 3480 | 4 | 245802107 | 6,266500 |
| 1562 | 2 | 363489782 | 27,80270356 | 1562 | 2 | 363489782 | 6,8718003 | 1562 | 2 | 363489782 | 5,390297 |
| 4317 | 5 | 204516414 | 10,48457176 | 4317 | 5 | 204516414 | 2,9194618 | 3261 | 4 | 212893509 | 4,081755 |
| 1093 | 2 | 20954874 | 5,7664066 | 1093 | 2 | 20954874 | 1,7868704 | 5673 | 6 | 331710150 | 3,983664 |
| 3261 | 4 | 212893509 | 5,38687284 | 3261 | 4 | 212893509 | 1,6927437 | 3517 | 4 | 245955842 | 3,790299 |
| 3517 | 4 | 245955842 | 4,39474071 | 3517 | 4 | 245955842 | 1,4430954 | 5846 | 6 | 393323235 | 3,772569 |
| 2749 | 3 | 596058675 | 2,22345874 | 2749 | 3 | 596058675 | 0,8666903 | 2517 | 3 | 380243227 | 3,400196 |
| 2517 | 3 | 380243227 | 2,10599434 | 2517 | 3 | 380243227 | 0,8335020 | 5933 | 6 | 424307850 | 3,338913 |
| 1417 | 2 | 259874800 | 1,62703882 | 1417 | 2 | 259874800 | 0,6944055 | 5277 | 6 | 190676730 | 3,331761 |
| 5673 | 6 | 331710150 | 1,51328894 | 5673 | 6 | 331710150 | 0,6602745 | 1646 | 2 | 423515013 | 3,277031 |
| 3960 | 4 | 247796507 | 1,32338858 | 3960 | 4 | 247796507 | 0,6020864 | 3736 | 4 | 246865787 | 3,273612 |
| 2346 | 3 | 223513366 | 1,13911370 | 2346 | 3 | 223513366 | 0,5438775 | 5884 | 6 | 406856745 | 3,248133 |
| 5461 | 6 | 256207410 | 0,99305717 | 5461 | 6 | 256207410 | 0,4962143 | 4317 | 5 | 204516414 | 3,192595 |
| 3971 | 4 | 247842212 | 0,98869622 | 3971 | 4 | 247842212 | 0,4947669 | 4649 | 5 | 486860350 | 3,177421 |
| 4913 | 5 | 731357614 | 0,83622657 | 4913 | 5 | 731357614 | 0,4431210 | 2749 | 3 | 596058675 | 3,177365 |
| 499 | 1 | 291031959 | 0,76783281 | 499 | 1 | 291031959 | 0,4192023 | 5862 | 6 | 399021555 | 3,117959 |
| 4987 | 5 | 799890938 | 0,74511608 | 4987 | 5 | 799890938 | 0,4111381 | 3381 | 4 | 245390762 | 3,035844 |
| 4549 | 5 | 394247750 | 0,67759552 | 4549 | 5 | 394247750 | 0,3867758 | 5670 | 6 | 330641715 | 3,008389 |
| 5577 | 6 | 297520230 | 0,67264696 | 5577 | 6 | 297520230 | 0,3849655 | 187 | 1 | 83595286 | 2,927185 |
| 1252 | 2 | 164561951 | 0,66246706 | 1252 | 2 | 164561951 | 0,3812301 | 5799 | 6 | 376584420 | 2,906733 |
| 945 | 1 | 614932841 | 0,61656550 | 945 | 1 | 614932841 | 0,3641882 | 1093 | 2 | 20954874 | 2,879809 |
| 187 | 1 | 83595286 | 0,57007201 | 187 | 1 | 83595286 | 0,3465653 | 1865 | 2 | 580009364 | 2,858210 |
| 2602 | 3 | 459313542 | 0,55444660 | 2602 | 3 | 459313542 | 0,3405530 | 3389 | 4 | 245424002 | 2,708215 |
| 1646 | 2 | 423515013 | 0,51218800 | 1646 | 2 | 423515013 | 0,3240447 | 5033 | 6 | 75187712 | 2,700129 |
| 1095 | 2 | 20958307 | 0,49476658 | 1095 | 2 | 20958307 | 0,3171254 | 5805 | 6 | 378721290 | 2,686021 |
| 4629 | 5 | 468337830 | 0,45683825 | 4629 | 5 | 468337830 | 0,3018078 | 427 | 1 | 268373293 | 2,590842 |
| 4081 | 5 | 19973330 | 0,43593349 | 4081 | 5 | 19973330 | 0,2932039 | 4290 | 5 | 193521948 | 2,568202 |
| 4138 | 5 | 62622237 | 0,42992716 | 4138 | 5 | 62622237 | 0,2907091 | 945 | 1 | 614932841 | 2,566545 |
| 514 | 1 | 293632633 | 0,36674051 | 514 | 1 | 293632633 | 0,2637738 | 5698 | 6 | 340613775 | 2,563206 |
| 4554 | 5 | 398878380 | 0,34594581 | 4554 | 5 | 398878380 | 0,2545995 | 4265 | 5 | 184588958 | 2,491110 |
| 2088 | 3 | 69892547 | 0,33566283 | 2088 | 3 | 69892547 | 0,2499983 | 4838 | 5 | 661898164 | 2,438240 |
| 1230 | 2 | 144528591 | 0,33374192 | 1230 | 2 | 144528591 | 0,2491338 | 4453 | 5 | 305339654 | 2,422658 |
| 4337 | 5 | 209498142 | 0,33067894 | 4337 | 5 | 209498142 | 0,2477521 | 1095 | 2 | 20958307 | 2,421069 |
| 4170 | 5 | 75234294 | 0,32911169 | 4170 | 5 | 75234294 | 0,2470435 | 3133 | 4 | 128632208 | 2,401665 |
| 1534 | 2 | 343481372 | 0,32437646 | 1534 | 2 | 343481372 | 0,2448959 | 3096 | 4 | 76123039 | 2,394804 |
| 4341 | 5 | 210890745 | 0,30968297 | 4341 | 5 | 210890745 | 0,2381662 | 5619 | 6 | 312478320 | 2,369837 |
| 3315 | 4 | 245051599 | 0,30277873 | 3315 | 4 | 245051599 | 0,2349684 | 2877 | 3 | 715129267 | 2,355939 |
| 970 | 1 | 631685492 | 0,30189581 | 970 | 1 | 631685492 | 0,2345577 | 3971 | 4 | 247842212 | 2,332747 |
| 2800 | 3 | 643500864 | 0,30049455 | 2800 | 3 | 643500864 | 0,2339051 | 4802 | 5 | 628557628 | 2,332568 |
| 4648 | 5 | 485934224 | 0,29870256 | 4648 | 5 | 485934224 | 0,2330692 | 4648 | 5 | 485934224 | 2,326017 |
| 1263 | 2 | 169265278 | 0,29573939 | 1263 | 2 | 169265278 | 0,2316832 | 5244 | 6 | 178923945 | 2,324638 |
| 5846 | 6 | 393323235 | 0,29155465 | 5846 | 6 | 393323235 | 0,2297181 | 1358 | 2 | 214423504 | 2,304391 |
| 2874 | 3 | 712338550 | 0,28624607 | 2874 | 3 | 712338550 | 0,2272118 | 3972 | 4 | 247846367 | 2,286949 |
| 3773 | 4 | 247019522 | 0,28322354 | 3773 | 4 | 247019522 | 0,2257779 | 499 | 1 | 291031959 | 2,284529 |
| 2245 | 3 | 185290309 | 0,28085651 | 2245 | 3 | 185290309 | 0,2246515 | 1185 | 2 | 58354695 | 2,283718 |
| 5794 | 6 | 374803695 | 0,28001618 | 5794 | 6 | 374803695 | 0,2242508 | 5103 | 6 | 113384656 | 2,265262 |
| 4838 | 5 | 661898164 | 0,27711045 | 4838 | 5 | 661898164 | 0,2228623 | 1321 | 2 | 198757428 | 2,264077 |
| 5724 | 6 | 349873545 | 0,25481957 | 5724 | 6 | 349873545 | 0,2120427 | 821 | 1 | 531839692 | 2,263088 |
| 43 | 1 | 10067746 | 0,24258773 | 43 | 1 | 10067746 | 0,2059700 | 308 | 1 | 216989434 | 2,252107 |
| 4453 | 5 | 305339654 | 0,23845969 | 4453 | 5 | 305339654 | 0,2038971 | 1354 | 2 | 214263400 | 2,231001 |
| 1354 | 2 | 214263400 | 0,23609455 | 1354 | 2 | 214263400 | 0,2027039 | 5547 | 6 | 286835880 | 2,222143 |
| 2746 | 3 | 593267958 | 0,23158747 | 2746 | 3 | 593267958 | 0,2004188 | 4629 | 5 | 468337830 | 2,186507 |
| 1167 | 2 | 44606542 | 0,23140231 | 1167 | 2 | 44606542 | 0,2003246 | 4170 | 5 | 75234294 | 2,178396 |
| 1653 | 2 | 428517115 | 0,22852374 | 1653 | 2 | 428517115 | 0,1988567 | 3992 | 4 | 247929467 | 2,174125 |
| 1295 | 2 | 186512163 | 0,22549761 | 1295 | 2 | 186512163 | 0,1973066 | 3005 | 4 | 2637306 | 2,168391 |
| 1319 | 2 | 198756269 | 0,22087941 | 1319 | 2 | 198756269 | 0,1949271 | 2072 | 3 | 53549251 | 2,162393 |
| 1865 | 2 | 580009364 | 0,22061614 | 1865 | 2 | 580009364 | 0,1947909 | 1417 | 2 | 259874800 | 2,154835 |
| 5670 | 6 | 330641715 | 0,21646934 | 5670 | 6 | 330641715 | 0,1926386 | 5789 | 6 | 373022970 | 2,123864 |
| 5862 | 6 | 399021555 | 0,21276249 | 5862 | 6 | 399021555 | 0,1907025 | 1561 | 2 | 362775196 | 2,111705 |

Fonte: Da autora (2019)

Tabela 49 – BSSV modificado × GWAS
cenário Poligênico, $h^2 = 0,8$

| BSSV modificado | | | | | | | | GWAS | | | |
|-----------------|----|-----------|------------|---|----|-----------|--------------------------------|--------------------------------|----|-----------|--------------------------------|
| wald | | | | -log ₁₀ (valor - p) teste Wald | | | | -log ₁₀ (valor - p) | | | |
| Marca | Cr | Pos | wald | Marca | Cr | Pos | -log ₁₀ (valor - p) | Marca | Cr | Pos | -log ₁₀ (valor - p) |
| 5826 | 6 | 386200335 | 91,5793352 | 5826 | 6 | 386200335 | 20,9697779 | 5826 | 6 | 386200335 | 11,587916 |
| 1562 | 2 | 363489782 | 57,4943463 | 1562 | 2 | 363489782 | 13,4698616 | 3480 | 4 | 245802107 | 8,318912 |
| 3480 | 4 | 245802107 | 55,3262891 | 3480 | 4 | 245802107 | 12,9909998 | 5954 | 6 | 431786895 | 5,722483 |
| 4629 | 5 | 468337830 | 20,3262126 | 4629 | 5 | 468337830 | 5,1850879 | 1562 | 2 | 363489782 | 5,662491 |
| 2517 | 3 | 380243227 | 20,0726258 | 2517 | 3 | 380243227 | 5,1275161 | 5033 | 6 | 75187712 | 5,350154 |
| 4549 | 5 | 394247750 | 19,9273471 | 4549 | 5 | 394247750 | 5,0945203 | 5862 | 6 | 399021555 | 4,623745 |
| 1093 | 2 | 20954874 | 18,7835183 | 1093 | 2 | 20954874 | 4,8343786 | 2517 | 3 | 380243227 | 4,065757 |
| 3517 | 4 | 245955842 | 15,0879421 | 3517 | 4 | 245955842 | 3,9887821 | 5590 | 6 | 302150115 | 3,514904 |
| 3261 | 4 | 212893509 | 11,0139542 | 3261 | 4 | 212893509 | 3,0436945 | 3736 | 4 | 246865787 | 3,498160 |
| 3960 | 4 | 247796507 | 7,8212684 | 3960 | 4 | 247796507 | 2,2870563 | 1093 | 2 | 20954874 | 3,365941 |
| 3288 | 4 | 235779015 | 6,9727474 | 3288 | 4 | 235779015 | 2,0821778 | 2292 | 3 | 200201258 | 3,072857 |
| 5033 | 6 | 75187712 | 6,7184694 | 5033 | 6 | 75187712 | 2,0203629 | 3517 | 4 | 245955842 | 3,015800 |
| 4951 | 5 | 766550402 | 4,5922709 | 4951 | 5 | 766550402 | 1,4932727 | 859 | 1 | 557303721 | 2,967605 |
| 3318 | 4 | 245131700 | 3,9363099 | 3318 | 4 | 245131700 | 1,3255584 | 3643 | 4 | 246479372 | 2,908547 |
| 5862 | 6 | 399021555 | 3,4446189 | 5862 | 6 | 399021555 | 1,1975132 | 3996 | 4 | 247946087 | 2,879696 |
| 5560 | 6 | 291465765 | 2,8353204 | 5560 | 6 | 291465765 | 1,0352106 | 89 | 1 | 28084996 | 2,879099 |
| 3476 | 4 | 245785487 | 2,2956906 | 3476 | 4 | 245785487 | 0,8869479 | 3918 | 4 | 247621997 | 2,809284 |
| 89 | 1 | 28084996 | 1,7847945 | 89 | 1 | 28084996 | 0,7409758 | 5134 | 6 | 131576099 | 2,790126 |
| 5590 | 6 | 302150115 | 1,7643340 | 5590 | 6 | 302150115 | 0,7349820 | 1865 | 2 | 580009364 | 2,786465 |
| 3488 | 4 | 245835347 | 1,6722878 | 3488 | 4 | 245835347 | 0,7078495 | 4629 | 5 | 468337830 | 2,781031 |
| 382 | 1 | 262715518 | 1,6236437 | 382 | 1 | 262715518 | 0,6933938 | 5636 | 6 | 318532785 | 2,764711 |
| 4085 | 5 | 20200987 | 1,3330776 | 4085 | 5 | 20200987 | 0,6050960 | 1304 | 2 | 189293647 | 2,700383 |
| 3406 | 4 | 245494637 | 1,0622935 | 3406 | 4 | 245494637 | 0,5189981 | 5186 | 6 | 157129197 | 2,698615 |
| 5643 | 6 | 321025800 | 1,0469089 | 5643 | 6 | 321025800 | 0,5139663 | 5063 | 6 | 83064740 | 2,698013 |
| 4648 | 5 | 485934224 | 1,0369580 | 4648 | 5 | 485934224 | 0,5107026 | 5805 | 6 | 378721290 | 2,687730 |
| 5109 | 6 | 113385418 | 1,0151673 | 5109 | 6 | 113385418 | 0,5035297 | 3476 | 4 | 245785487 | 2,630147 |
| 5472 | 6 | 260125005 | 1,0126929 | 5472 | 6 | 260125005 | 0,5027129 | 1646 | 2 | 423515013 | 2,593889 |
| 3337 | 4 | 245207942 | 0,959965 | 3337 | 4 | 245207942 | 0,4851896 | 5801 | 6 | 377296710 | 2,582850 |
| 1528 | 2 | 339193855 | 0,9197655 | 1528 | 2 | 339193855 | 0,4716792 | 2840 | 3 | 680710424 | 2,569747 |
| 4849 | 5 | 672085550 | 0,8966406 | 4849 | 5 | 672085550 | 0,4638406 | 966 | 1 | 629005068 | 2,547537 |
| 1686 | 2 | 452098456 | 0,7782196 | 1686 | 2 | 452098456 | 0,4228689 | 5530 | 6 | 280781415 | 2,532024 |
| 5461 | 6 | 256207410 | 0,6847117 | 5461 | 6 | 256207410 | 0,3893729 | 4459 | 5 | 310896410 | 2,531509 |
| 2289 | 3 | 200142461 | 0,6639598 | 2289 | 3 | 200142461 | 0,3817788 | 5416 | 6 | 240180885 | 2,488117 |
| 3416 | 4 | 245536187 | 0,6605717 | 3416 | 4 | 245536187 | 0,3805329 | 1933 | 2 | 628601218 | 2,436105 |
| 4593 | 5 | 434997294 | 0,6528092 | 4593 | 5 | 434997294 | 0,3776719 | 3658 | 4 | 246541697 | 2,433635 |
| 158 | 1 | 56175806 | 0,6111566 | 158 | 1 | 56175806 | 0,3621575 | 3538 | 4 | 246043097 | 2,428786 |
| 5416 | 6 | 240180885 | 0,6094368 | 5416 | 6 | 240180885 | 0,3615108 | 5442 | 6 | 249440655 | 2,404876 |
| 1249 | 2 | 161879250 | 0,5837908 | 1249 | 2 | 161879250 | 0,3518057 | 3220 | 4 | 180309373 | 2,391637 |
| 3032 | 4 | 28192994 | 0,5816157 | 3032 | 4 | 28192994 | 0,3509771 | 5290 | 6 | 195306615 | 2,384984 |
| 3448 | 4 | 245669147 | 0,5793838 | 3448 | 4 | 245669147 | 0,3501261 | 382 | 1 | 262715518 | 2,379969 |
| 2279 | 3 | 194643731 | 0,5762338 | 2279 | 3 | 194643731 | 0,3489233 | 5782 | 6 | 370529955 | 2,373650 |
| 5131 | 6 | 129927781 | 0,5614156 | 5131 | 6 | 129927781 | 0,3432404 | 1086 | 2 | 19837545 | 2,371760 |
| 3313 | 4 | 244992633 | 0,5122228 | 3313 | 4 | 244992633 | 0,3240584 | 4549 | 5 | 394247750 | 2,352150 |
| 5417 | 6 | 240537030 | 0,5102917 | 5417 | 6 | 240537030 | 0,3232949 | 2998 | 3 | 827688186 | 2,352145 |
| 1646 | 2 | 423515013 | 0,4967524 | 1646 | 2 | 423515013 | 0,3179176 | 1880 | 2 | 590728155 | 2,342969 |
| 3220 | 4 | 180309373 | 0,4838581 | 3220 | 4 | 180309373 | 0,3127566 | 5698 | 6 | 340613775 | 2,338420 |
| 2901 | 3 | 737455003 | 0,4022933 | 2901 | 3 | 737455003 | 0,2790908 | 3295 | 4 | 238553534 | 2,331521 |
| 4940 | 5 | 756363016 | 0,3956033 | 4940 | 5 | 756363016 | 0,2762418 | 5171 | 6 | 149251173 | 2,318388 |
| 3576 | 4 | 246200987 | 0,3904524 | 3576 | 4 | 246200987 | 0,2740380 | 3337 | 4 | 245207942 | 2,312483 |
| 126 | 1 | 44691505 | 0,3893559 | 126 | 1 | 44691505 | 0,2735677 | 728 | 1 | 469519830 | 2,297564 |
| 3755 | 4 | 246944732 | 0,3807464 | 3755 | 4 | 246944732 | 0,2698607 | 1686 | 2 | 452098456 | 2,294555 |
| 5636 | 6 | 318532785 | 0,3729870 | 5636 | 6 | 318532785 | 0,2664973 | 3632 | 4 | 246433667 | 2,263630 |
| 1087 | 2 | 20042233 | 0,3729077 | 1087 | 2 | 20042233 | 0,2664629 | 3261 | 4 | 212893509 | 2,223022 |
| 3679 | 4 | 246628952 | 0,3705560 | 3679 | 4 | 246628952 | 0,2654391 | 5641 | 6 | 320313510 | 2,207166 |
| 251 | 1 | 183647544 | 0,3630139 | 251 | 1 | 183647544 | 0,2621420 | 5244 | 6 | 178923945 | 2,205616 |
| 3632 | 4 | 246433667 | 0,3566685 | 3632 | 4 | 246433667 | 0,2593513 | 4849 | 5 | 672085550 | 2,204546 |
| 3663 | 4 | 246562472 | 0,3233639 | 3663 | 4 | 246562472 | 0,2444353 | 1249 | 2 | 161879250 | 2,186632 |
| 5380 | 6 | 227359665 | 0,3224882 | 5380 | 6 | 227359665 | 0,2440367 | 5336 | 6 | 211689285 | 2,170129 |
| 1902 | 2 | 606449049 | 0,3206784 | 1902 | 2 | 606449049 | 0,2432116 | 4394 | 5 | 250698220 | 2,163688 |
| 325 | 1 | 224518409 | 0,3148152 | 325 | 1 | 224518409 | 0,2405284 | 1867 | 2 | 581438536 | 2,150772 |

Fonte: Da autora (2019)

Tabela 50 – FASTmrEMMA
cenário Oligogênico

| herdabilidade 0,2 | | | | herdabilidade 0,5 | | | | herdabilidade 0,8 | | | |
|-------------------|----|-----------|--------------------------------|-------------------|----|-----------|--------------------------------|-------------------|----|-----------|--------------------------------|
| Marca | Cr | Pos | $-\log_{10}(\text{valor} - p)$ | Marca | Cr | Pos | $-\log_{10}(\text{valor} - p)$ | Marca | Cr | Pos | $-\log_{10}(\text{valor} - p)$ |
| 4415 | 5 | 270146866 | 3,4191 | 258 | 1 | 188048080 | 2,4057 | 1238 | 2 | 157152675 | 2,6059 |
| 4065 | 5 | 12338040 | 2,9914 | 4463 | 5 | 314600914 | 2,3507 | 898 | 1 | 583437857 | 2,2808 |
| 4705 | 5 | 538723406 | 2,7870 | 5867 | 6 | 400802280 | 2,3234 | 5988 | 6 | 443895825 | 2,2641 |
| 2018 | 3 | 6733476 | 2,7239 | 5387 | 6 | 229852680 | 2,3225 | 3815 | 4 | 247194032 | 2,2302 |
| 4425 | 5 | 279408126 | 2,5932 | 5087 | 6 | 92705503 | 2,3034 | 338 | 1 | 233401162 | 2,1142 |
| 4035 | 5 | 7103536 | 2,5366 | 598 | 1 | 382406045 | 2,2951 | 5268 | 6 | 187471425 | 2,0553 |
| 5625 | 6 | 314615190 | 2,4975 | 4223 | 5 | 167884630 | 2,2446 | 678 | 1 | 436014528 | 2,0520 |
| 4195 | 5 | 100227859 | 2,4576 | 38 | 1 | 10065344 | 2,2241 | 1578 | 2 | 374923159 | 2,0336 |
| 4205 | 5 | 145121333 | 2,4193 | 1158 | 2 | 42722721 | 2,1690 | 2375 | 3 | 248149289 | 2,0217 |
| 865 | 1 | 561324358 | 2,2260 | 4997 | 5 | 809152198 | 2,1650 | 1018 | 2 | 2793372 | 2,0204 |

Fonte: Da autora (2019)

Tabela 51 – FASTmrEMMA
cenário Poligênico

| herdabilidade 0,2 | | | | herdabilidade 0,5 | | | | herdabilidade 0,8 | | | |
|-------------------|----|-----------|--------------------------------|-------------------|----|-----------|--------------------------------|-------------------|----|-----------|--------------------------------|
| Marca | Cr | Pos | $-\log_{10}(\text{valor} - p)$ | Marca | Cr | Pos | $-\log_{10}(\text{valor} - p)$ | Marca | Cr | Pos | $-\log_{10}(\text{valor} - p)$ |
| 4137 | 5 | 62415995 | 2,5489 | 1167 | 2 | 44606542 | 2,8694 | 4098 | 5 | 30239292 | 2,5729 |
| 4997 | 5 | 809152198 | 2,1856 | 1620 | 2 | 404935774 | 2,8293 | 3262 | 4 | 215393158 | 2,3706 |
| 5379 | 6 | 227003520 | 1,9372 | 497 | 1 | 290199645 | 2,5458 | 1620 | 2 | 404935774 | 2,3037 |
| 5389 | 6 | 230564970 | 1,9021 | 1777 | 2 | 517125789 | 2,5430 | 1937 | 2 | 631459562 | 2,1493 |
| 5348 | 6 | 215963025 | 1,8828 | 1797 | 2 | 531417511 | 2,4601 | 2182 | 3 | 162187230 | 2,1076 |
| 4559 | 5 | 403509010 | 1,8654 | 1937 | 2 | 631459562 | 2,4248 | 3322 | 4 | 245145617 | 2,0451 |
| 2520 | 3 | 383033944 | 1,8390 | 1667 | 2 | 438521320 | 2,4024 | 3462 | 4 | 245727317 | 2,0045 |
| 2446 | 3 | 314196258 | 1,8103 | 977 | 1 | 636376234 | 2,3891 | 1644 | 2 | 422085840 | 1,9874 |
| 3177 | 4 | 156290035 | 1,7767 | 1647 | 2 | 424229599 | 2,1946 | 2582 | 3 | 440708762 | 1,9676 |
| 4799 | 5 | 625779250 | 1,7602 | 1876 | 2 | 587869811 | 2,0605 | 1777 | 2 | 517125789 | 1,9099 |
| 2606 | 3 | 463034498 | 1,7590 | 1567 | 2 | 367062712 | 2,0257 | 2382 | 3 | 254660962 | 1,8705 |
| 4357 | 5 | 216431558 | 1,7578 | 1130 | 2 | 27790401 | 1,9365 | 4712 | 5 | 545206288 | 1,8509 |
| 3246 | 4 | 202883397 | 1,6923 | 1587 | 2 | 381354434 | 1,9354 | 1955 | 2 | 644322111 | 1,7962 |
| 4777 | 5 | 605404478 | 1,6821 | 1450 | 2 | 283456141 | 1,8861 | 5258 | 6 | 183909975 | 1,7962 |
| 1980 | 2 | 662186763 | 1,6740 | 1817 | 2 | 545709232 | 1,8791 | 2442 | 3 | 310475302 | 1,7856 |
| 2906 | 3 | 742106198 | 1,6649 | 5199 | 6 | 161663671 | 1,8675 | 1797 | 2 | 531417511 | 1,7814 |
| 4949 | 5 | 764698150 | 1,6645 | 5081 | 6 | 89099054 | 1,8661 | 3085 | 4 | 73311191 | 1,7750 |
| 1439 | 2 | 275595694 | 1,6587 | 667 | 1 | 428643362 | 1,8561 | 2272 | 3 | 193565631 | 1,7696 |
| 4778 | 5 | 606330604 | 1,6585 | 1007 | 2 | 1220908 | 1,8170 | 3862 | 4 | 247389317 | 1,7245 |
| 4019 | 5 | 5267091 | 1,6221 | 5281 | 6 | 192101310 | 1,7914 | 3362 | 4 | 245311817 | 1,7235 |
| 5003 | 6 | 6321890 | 1,6148 | 1690 | 2 | 454956800 | 1,7573 | 754 | 1 | 486942587 | 1,6925 |
| 5657 | 6 | 326011830 | 1,5851 | 4415 | 5 | 270146866 | 1,7462 | 115 | 1 | 41390404 | 1,6724 |
| 4838 | 5 | 661898164 | 1,5773 | 1613 | 2 | 399933672 | 1,7382 | 2129 | 3 | 133480594 | 1,6636 |
| 5589 | 6 | 301793970 | 1,5574 | 5141 | 6 | 135105656 | 1,7131 | 2342 | 3 | 222114627 | 1,6541 |
| 5897 | 6 | 411486630 | 1,5378 | 5578 | 6 | 297876375 | 1,7095 | 4852 | 5 | 674863928 | 1,6508 |
| 269 | 1 | 201295773 | 1,5214 | 4835 | 5 | 659119786 | 1,6819 | 554 | 1 | 352921379 | 1,6311 |
| 4765 | 5 | 594290966 | 1,4764 | 629 | 1 | 403179332 | 1,6452 | 3982 | 4 | 247887917 | 1,5999 |
| 229 | 1 | 161072169 | 1,4719 | 1909 | 2 | 611451152 | 1,6390 | 2809 | 3 | 651873015 | 1,5982 |
| 4868 | 5 | 689681944 | 1,4554 | 787 | 1 | 509056086 | 1,6389 | 1167 | 2 | 44606542 | 1,5861 |
| 4599 | 5 | 440554050 | 1,4553 | 469 | 1 | 285421542 | 1,6386 | 3302 | 4 | 242019759 | 1,5848 |
| 4850 | 5 | 673011676 | 1,4528 | 137 | 1 | 46810810 | 1,6235 | 2482 | 3 | 347684862 | 1,5759 |
| 5360 | 6 | 220236765 | 1,4501 | 1940 | 2 | 633603320 | 1,6147 | 2778 | 3 | 623035606 | 1,5649 |
| 5835 | 6 | 389405640 | 1,4427 | 917 | 1 | 596169872 | 1,6126 | 3512 | 4 | 245935067 | 1,5537 |
| 3397 | 4 | 245457242 | 1,4408 | 457 | 1 | 284407382 | 1,6057 | 2422 | 3 | 291870522 | 1,5485 |
| 4880 | 5 | 700795456 | 1,4407 | 4765 | 5 | 594290966 | 1,6021 | 4228 | 5 | 168867197 | 1,5443 |
| 804 | 1 | 520447889 | 1,4298 | 3335 | 4 | 245199632 | 1,5989 | 2274 | 3 | 193566873 | 1,5440 |
| 782 | 1 | 505705556 | 1,4288 | 1889 | 2 | 597159430 | 1,5974 | 3392 | 4 | 245436467 | 1,5437 |
| 1729 | 2 | 482825657 | 1,4104 | 269 | 1 | 201295773 | 1,5807 | 3102 | 4 | 86798226 | 1,5289 |
| 5779 | 6 | 369461520 | 1,4038 | 69 | 1 | 17900705 | 1,5622 | 3279 | 4 | 222331993 | 1,5281 |
| 5757 | 6 | 361626330 | 1,3967 | 1017 | 2 | 2527442 | 1,5517 | 4632 | 5 | 471116208 | 1,5182 |
| 4940 | 5 | 756363016 | 1,3962 | 1947 | 2 | 638605423 | 1,5517 | 615 | 1 | 393797847 | 1,5173 |
| 1189 | 2 | 59600932 | 1,3781 | 489 | 1 | 287179850 | 1,5496 | 84 | 1 | 26283149 | 1,5149 |
| 5029 | 6 | 57774504 | 1,3745 | 4971 | 5 | 785072922 | 1,5147 | 2242 | 3 | 184718109 | 1,5090 |
| 1769 | 2 | 511409100 | 1,3744 | 100 | 1 | 34479054 | 1,5140 | 1115 | 2 | 25924769 | 1,5040 |
| 1620 | 2 | 404935774 | 1,3655 | 376 | 1 | 260688178 | 1,5134 | 3222 | 4 | 181809360 | 1,4991 |
| 2017 | 3 | 6547051 | 1,3574 | 1307 | 2 | 193285152 | 1,5086 | 1193 | 2 | 62870552 | 1,4757 |

Fonte: Da autora (2019)

Tabela 52 – FASTmrMLM
cenário Oligogênico

| herdabilidade 0,2 | | | | herdabilidade 0,5 | | | | herdabilidade 0,8 | | | |
|-------------------|----|-----------|--------------------------------|-------------------|----|-----------|--------------------------------|-------------------|----|-----------|--------------------------------|
| Marca | Cr | Pos | $-\log_{10}(\text{valor} - p)$ | Marca | Cr | Pos | $-\log_{10}(\text{valor} - p)$ | Marca | Cr | Pos | $-\log_{10}(\text{valor} - p)$ |
| 4415 | 5 | 270146866 | 3,792085 | 258 | 1 | 188048080 | 2,703444 | 1238 | 2 | 157152675 | 2,887207 |
| 4065 | 5 | 12338040 | 3,335017 | 4463 | 5 | 314600914 | 2,652535 | 898 | 1 | 583437857 | 2,528544 |
| 4705 | 5 | 538723406 | 3,115332 | 5867 | 6 | 400802280 | 2,641849 | 5988 | 6 | 443895825 | 2,527707 |
| 2018 | 3 | 6733476 | 3,036050 | 5387 | 6 | 229852680 | 2,618047 | 3815 | 4 | 247194032 | 2,523479 |
| 4425 | 5 | 279408126 | 2,891724 | 5087 | 6 | 92705503 | 2,606708 | 338 | 1 | 233401162 | 2,342622 |
| 4035 | 5 | 7103536 | 2,828852 | 598 | 1 | 382406045 | 2,587564 | 5268 | 6 | 187471425 | 2,310159 |
| 5625 | 6 | 314615190 | 2,789743 | 4223 | 5 | 167884630 | 2,557462 | 2375 | 3 | 248149289 | 2,300313 |
| 4195 | 5 | 100227859 | 2,746908 | 38 | 1 | 10065344 | 2,504827 | 1578 | 2 | 374923159 | 2,281944 |
| 4205 | 5 | 145121333 | 2,700974 | 1158 | 2 | 42722721 | 2,469910 | 678 | 1 | 436014528 | 2,281381 |
| 865 | 1 | 561324358 | 2,487268 | 4997 | 5 | 809152198 | 2,436106 | 5448 | 6 | 251577525 | 2,279500 |

Fonte: Da autora (2019)

Tabela 53 – FASTmrMLM
cenário Poligênico

| herdabilidade 0,2 | | | | herdabilidade 0,5 | | | | herdabilidade 0,8 | | | |
|-------------------|----|-----------|--------------------------------|-------------------|----|-----------|--------------------------------|-------------------|----|-----------|--------------------------------|
| Marca | Cr | Pos | $-\log_{10}(\text{valor} - p)$ | Marca | Cr | Pos | $-\log_{10}(\text{valor} - p)$ | Marca | Cr | Pos | $-\log_{10}(\text{valor} - p)$ |
| 1620 | 2 | 404935774 | 1,521400 | 1167 | 2 | 44606542 | 3,192895 | 1620 | 2 | 404935774 | 2,537129 |
| 2017 | 3 | 6547051 | 1,513151 | 1620 | 2 | 404935774 | 3,148503 | 2182 | 3 | 162187230 | 2,424287 |
| 1532 | 2 | 342052199 | 1,506348 | 497 | 1 | 290199645 | 2,839019 | 1937 | 2 | 631459562 | 2,414421 |
| 649 | 1 | 416581453 | 1,488477 | 1777 | 2 | 517125789 | 2,836083 | 3322 | 4 | 245145617 | 2,333027 |
| 5529 | 6 | 280425270 | 1,486456 | 1797 | 2 | 531417511 | 2,749391 | 3462 | 4 | 245727317 | 2,288813 |
| 5189 | 6 | 159808180 | 1,481349 | 1937 | 2 | 631459562 | 2,708183 | 2582 | 3 | 440708762 | 2,254321 |
| 4852 | 5 | 674863928 | 1,479930 | 1667 | 2 | 438521320 | 2,688734 | 1644 | 2 | 422085840 | 2,234402 |
| 4903 | 5 | 722096354 | 1,474433 | 977 | 1 | 636376234 | 2,667044 | 1777 | 2 | 517125789 | 2,202918 |
| 4318 | 5 | 204516488 | 1,471460 | 1647 | 2 | 424229599 | 2,452724 | 1955 | 2 | 644322111 | 2,123129 |
| 4834 | 5 | 658193660 | 1,469145 | 1876 | 2 | 587869811 | 2,305764 | 2382 | 3 | 254660962 | 2,119669 |
| 1100 | 2 | 21202350 | 1,461599 | 1567 | 2 | 367062712 | 2,266381 | 5258 | 6 | 183909975 | 2,104890 |
| 4973 | 5 | 786925174 | 1,461161 | 1130 | 2 | 27790401 | 2,180364 | 4712 | 5 | 545206288 | 2,099550 |
| 3321 | 4 | 245141462 | 1,456960 | 1587 | 2 | 381354434 | 2,168594 | 1797 | 2 | 531417511 | 2,077222 |
| 1835 | 2 | 558571782 | 1,447417 | 1450 | 2 | 283456141 | 2,111813 | 2442 | 3 | 310475302 | 2,041907 |
| 1209 | 2 | 125642617 | 1,447319 | 1817 | 2 | 545709232 | 2,109002 | 2272 | 3 | 193565631 | 2,032505 |
| 1389 | 2 | 233874044 | 1,446003 | 5199 | 6 | 161663671 | 2,102021 | 3085 | 4 | 73311191 | 1,981846 |
| 5533 | 6 | 281849850 | 1,443343 | 5081 | 6 | 89099054 | 2,096098 | 3362 | 4 | 245311817 | 1,977610 |
| 1932 | 2 | 627886631 | 1,435397 | 667 | 1 | 428643362 | 2,076931 | 115 | 1 | 41390404 | 1,976310 |
| 4419 | 5 | 273851370 | 1,431932 | 1007 | 2 | 1220908 | 2,033690 | 3862 | 4 | 247389317 | 1,965600 |
| 559 | 1 | 356271909 | 1,431688 | 5281 | 6 | 192101310 | 2,006406 | 2129 | 3 | 133480594 | 1,931238 |
| 3040 | 4 | 33860915 | 1,419810 | 1690 | 2 | 454956800 | 1,982917 | 754 | 1 | 486942587 | 1,931131 |
| 3798 | 4 | 247123397 | 1,414725 | 4415 | 5 | 270146866 | 1,955508 | 2342 | 3 | 222114627 | 1,898793 |
| 4383 | 5 | 240510834 | 1,410717 | 1613 | 2 | 399933672 | 1,945305 | 4852 | 5 | 674863928 | 1,878323 |
| 4085 | 5 | 20200987 | 1,409803 | 5141 | 6 | 135105656 | 1,917322 | 554 | 1 | 352921379 | 1,840228 |
| 5118 | 6 | 118741880 | 1,399365 | 5578 | 6 | 297876375 | 1,913892 | 3302 | 4 | 242019759 | 1,833984 |
| 1590 | 2 | 383498192 | 1,392078 | 4835 | 5 | 659119786 | 1,888201 | 3982 | 4 | 247887917 | 1,826897 |
| 829 | 1 | 537200540 | 1,391675 | 629 | 1 | 403179332 | 1,857904 | 1167 | 2 | 44606542 | 1,825531 |
| 4298 | 5 | 199531408 | 1,389734 | 469 | 1 | 285421542 | 1,846269 | 4228 | 5 | 168867197 | 1,817370 |
| 3080 | 4 | 73307596 | 1,387752 | 1909 | 2 | 611451152 | 1,844379 | 2778 | 3 | 623035606 | 1,816312 |
| 4379 | 5 | 236806330 | 1,387561 | 787 | 1 | 509056086 | 1,839401 | 2482 | 3 | 347684862 | 1,813349 |
| 5258 | 6 | 183909975 | 1,384983 | 137 | 1 | 46810810 | 1,815661 | 3512 | 4 | 245935067 | 1,810034 |
| 5519 | 6 | 276863820 | 1,384482 | 457 | 1 | 284407382 | 1,813263 | 2422 | 3 | 291870522 | 1,790800 |
| 5520 | 6 | 277219965 | 1,379666 | 1940 | 2 | 633603320 | 1,805892 | 2809 | 3 | 651873015 | 1,787595 |
| 1223 | 2 | 140014895 | 1,379396 | 917 | 1 | 596169872 | 1,801691 | 3392 | 4 | 245436467 | 1,775514 |
| 3076 | 4 | 73177100 | 1,376094 | 4765 | 5 | 594290966 | 1,792807 | 615 | 1 | 393797847 | 1,769341 |
| 436 | 1 | 275190900 | 1,373450 | 1889 | 2 | 597159430 | 1,790008 | 2242 | 3 | 184718109 | 1,764294 |
| 2441 | 3 | 309545063 | 1,365635 | 3335 | 4 | 245199632 | 1,788154 | 3279 | 4 | 222331993 | 1,761345 |
| 972 | 1 | 633025704 | 1,359966 | 269 | 1 | 201295773 | 1,769146 | 4632 | 5 | 471116208 | 1,756705 |
| 4429 | 5 | 283112630 | 1,356030 | 69 | 1 | 17900705 | 1,758035 | 2274 | 3 | 193566873 | 1,753562 |
| 2700 | 3 | 550476964 | 1,355565 | 1947 | 2 | 638605423 | 1,734336 | 3102 | 4 | 86798226 | 1,749369 |
| 1119 | 2 | 27403594 | 1,353148 | 1017 | 2 | 2527442 | 1,734290 | 4698 | 5 | 532240524 | 1,743638 |
| 1042 | 2 | 7473644 | 1,350042 | 489 | 1 | 287179850 | 1,732744 | 1115 | 2 | 25924769 | 1,735498 |
| 3406 | 4 | 245494637 | 1,345862 | 4971 | 5 | 785072922 | 1,709261 | 1433 | 2 | 271308177 | 1,733321 |
| 4578 | 5 | 421105404 | 1,340808 | 376 | 1 | 260688178 | 1,694578 | 84 | 1 | 26283149 | 1,732211 |
| 3337 | 4 | 245207942 | 1,340622 | 100 | 1 | 34479054 | 1,691623 | 3752 | 4 | 246932267 | 1,729492 |
| 1370 | 2 | 222280629 | 1,339141 | 1307 | 2 | 193285152 | 1,685737 | 1193 | 2 | 62870552 | 1,727762 |
| 3059 | 4 | 41191526 | 1,329713 | 507 | 1 | 291834351 | 1,678744 | 3222 | 4 | 181809360 | 1,723524 |

Fonte: Da autora (2019)

ANEXO D- USO DOS PACOTES PARA ANÁLISE PELO SOFTWARE R

Métodos RR-BLUP, Bayes A e Bayes B usando pacotes rrBLUP e BGLR

O pacote rrBLUP foi desenvolvido principalmente para predição genômica com modelos mistos (também executa análises GWAS). A função considerada de grande relevância é `mixed.solve`, que é uma função com propósito geral de solucionar análises de modelos mistos, com um único componente de variância diferente do erro (ENDELMAN, 2011).

A função `mixed.solve`, do pacote `rrBLUP`, pode ser usada para prever efeitos dos marcadores ou valores genéticos. O método é baseado na decomposição espectral de \mathbf{ZKZ}' e $\mathbf{SZKZ}'\mathbf{S}$, no qual $\mathbf{S} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$ (KANG et al., 2008). Este algoritmo gera a matriz de covariância fenotípica inversa \mathbf{V}^{-1} , que pode então ser usada para calcular as soluções BLUE e BLUP para os efeitos fixos e aleatórios, respectivamente, usando fórmulas padrão

$$BLUE(\boldsymbol{\beta}) = \hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{y} \quad (2)$$

$$BLUP(\mathbf{a}) = \hat{\mathbf{a}} = \sigma_a^2 \mathbf{KZ}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \quad (3)$$

o qual $\boldsymbol{\beta}$ é um vetor de efeitos fixos e \mathbf{a} é um vetor de efeitos aleatórios. Com $Var[\mathbf{a}] = \mathbf{K}\sigma_a^2$ em que \mathbf{K} é a matriz de covariância dos efeitos aleatórios. A variância residual é $Var[\mathbf{e}] = \mathbf{I}\sigma_e^2$. O erro padrão é calculado como a raiz quadrada dos elementos diagonais das seguintes matrizes:

$$Var[\hat{\boldsymbol{\beta}}] = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1} \quad (4)$$

$$Var[\hat{\mathbf{a}} - \mathbf{a}] = \mathbf{K}\sigma_a^2 - \sigma_a^4 \mathbf{KZ}'\mathbf{V}^{-1}\mathbf{ZK} + \sigma_a^4 \mathbf{KZ}'\mathbf{V}^{-1}\mathbf{X}Var[\hat{\boldsymbol{\beta}}]\mathbf{X}'\mathbf{V}^{-1}\mathbf{ZK} \quad (5)$$

Com uso da função `mixed.solve(y, Z=, K=, X=, method="REML", bounds=c(1e-09, 1e+09), SE=FALSE, return.Hinv=FALSE)`, retorna uma lista contendo valores como (σ_a^2) , σ_e^2 , $BLUE(\boldsymbol{\beta})$ e $BLUP(\mathbf{a})$, dentre outros. Outros detalhes em (ENDELMAN, 2011).

Sob outra perspectiva, o pacote BGLR (PÉREZ; CAMPOS, 2014) implementa uma variedade de procedimentos de regressão e de seleção de variáveis. Os métodos Bayes A e Bayes B foram implementados pelo pacote BGLR.

O modelo Bayes B (MEUWISSEN; HAYES; GODDARD, 2001), implementado pelo pacote BGLR usa priori para os efeitos que são misturas finitas, com um ponto de massa em zero e uma distribuição t-escalonada e para priori dos hiperparâmetros, considera-se a proporção de efeitos não nulos. Neste caso, com uso do BGLR pode-se atribuir uma distribuição a priori Beta para este hiperparâmetro. A versão da priori Beta passa a ser indexada por dois parâmetros de forma (`shape1` e `shape2` pelo *software* R) que podem ser considerados como o número de sucessos e fracassos em testes de priori de Bernoulli, rotulados como `counts0` e `counts1`, respectivamente. Por padrão, o BGLR define `counts0=counts1=5`. Para o método Bayes A, é usada a densidade t-escalonada Meuwissen, Hayes e Goddard (2001), com centro de massa maior em zero e caudas mais grossas que a densidade normal, que induz ao encolhimento de estimativas que dependem do tamanho do efeito (GIANOLA, 2013).

Cada distribuição descrita pelo BGLR é indexada por um ou mais parâmetros, que controlam o tipo e a extensão do encolhimento induzido. Detalhes sobre cada distribuição, tratada pelo pacote BGLR, pode ser vista a partir da tabela 55.

Tabela 54 – Distribuições a priori disponíveis para coeficientes de regressão no pacote BGLR

| Modelo (distribuição a priori) | Hiperparâmetros | Tratamento em BGLR |
|---------------------------------|---|--|
| t-escalonada Bayes A | graus de liberdade (df_{β}) escala (S_{β}) | $\mu_{\beta} = 0$ especificado pelo usuário (valor padrão, 5) |
| Mistura t-escalonada Bayes B | π (prob. de efeitos não nulos) df_{β} (S_{β}) | $\pi \sim \text{Beta}$ especificado pelo usuário (valor padrão, 5) $(S_{\beta}) \sim \text{Gama}$ |

Fonte: Adaptado (PÉREZ; CAMPOS, 2014)

Tabela 55 – Distribuições a priori implementadas no pacote BGLR

| model= | distribuição conjunta dos efeitos e hiperparâmetros | Especificação de elementos no preditor linear |
|--------|---|--|
| BayesA | $p(\beta_j, \sigma_{\beta_j}^2, S_{\beta}) = \left\{ \prod_k N(\beta_{jk} 0, \sigma_{\beta_{jk}}^2) \chi^{-2} \left(\sigma_{\beta_{jk}}^2 df_{\beta}, S_{\beta} \right) \right\} G(S_{\beta} r, s)$ | list(X=, model="BayesA", df0=, rate0=, shape0=, R2=) |
| BayesB | $p(\beta_j, \sigma_{\beta_j}^2, \pi) = \left\{ \left[\prod_k N(\beta_{jk} 0, \sigma_{\beta_j}^2) + (1 - \pi) 1(\beta_{jk} = 0) \right] \times \chi^{-2} \left(\sigma_{\beta_j}^2 df_{\beta}, S_{\beta} \right) B(\pi \rho_0, \pi_0) G(S_{\beta} r, s) \right\}$ | list(X=, model="BayesB", df0=, rate0=, shape0=, probIn=, counts=, R2=) |

Fonte: Adaptado (PÉREZ; CAMPOS, 2014)

Pela tabela 55, $N(\cdot | \cdot, \cdot)$, $\chi^{-2}(\cdot | \cdot, \cdot)$, $G(\cdot | \cdot, \cdot)$, $B(\cdot | \cdot, \cdot)$ indicam a Normal, Qui-quadrado invertida escalonada, Gama e Beta, respectivamente. O argumento `probIn` representa a probabilidade a priori de um marcador ter um efeito não-nulo π_0 , `counts` (o número de "contagens anteriores") pode ser usado para controlar o quão informativo é a priori.

GWAS com pacote rrBLUP

A função para mapeamento associativo é chamada GWAS, implementada pelo pacote rrBLUP (ENDELMAN, 2011) e executa análise de associação genômica ampla com base no modelo linear misto de Yu et al. (2006)

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{g} + \mathbf{S}\boldsymbol{\tau} + \mathbf{e} \quad (6)$$

em que $\boldsymbol{\beta}$ é um vetor de efeitos fixos, que podem modelar tanto fatores ambientais, quanto estrutura populacional. A variável \mathbf{g} modela o efeito genético de cada linha como um efeito

aleatório com $Var[g] = \mathbf{K}\sigma_g^2$. A variável τ modela o efeito aditivo do SNP como um efeito fixo. A variância residual é $Var[\mathbf{e}] = \mathbf{I}\sigma_e^2$. O uso da função GWAS pode ser declarado como os argumentos padrão:

```
GWAS(pheno, geno, fixed=NULL, K=NULL, n.PC=0,
min.MAF=0.05, n.core=1, P3D=TRUE, plot=TRUE)
```

Para ilustrar seu uso, faz-se o mapeamento de associação considerando dados hipotéticos composto de uma genotipagem de 10010 SNPs e 300 observações fenotipados. Ao digitar as linhas de comando no R, os arquivos dos dados fenotípicos (myY) e genotípicos (myG) são lidos:

```
myY <- read.table("fenotipo.txt", head=TRUE)
myG <- read.table("genotipo.txt", head=TRUE)
```

O arquivo myG contém informações dos SNPs, com as primeiras três colunas: identificação marca (marca), cromossomo num total de 5 (crom) e posição (pos), respectivamente. Para a função GWAS, o código da matriz de marcadores deve seguir o formato $\{-1,0,1\}$, necessário ao uso quando se tratando do pacote rrBLUP. A última linha de implementação mostra as dimensões do arquivo que contém a matriz de marcadores (desconsiderando as 3 colunas iniciais referente a informações dos SNPs): 10010 linhas por 300 observações.

```
> head(myY)
line      y
G1      1 -1.327212
G2      2  6.542220
G3      3  2.173966
G4      4 -7.579218

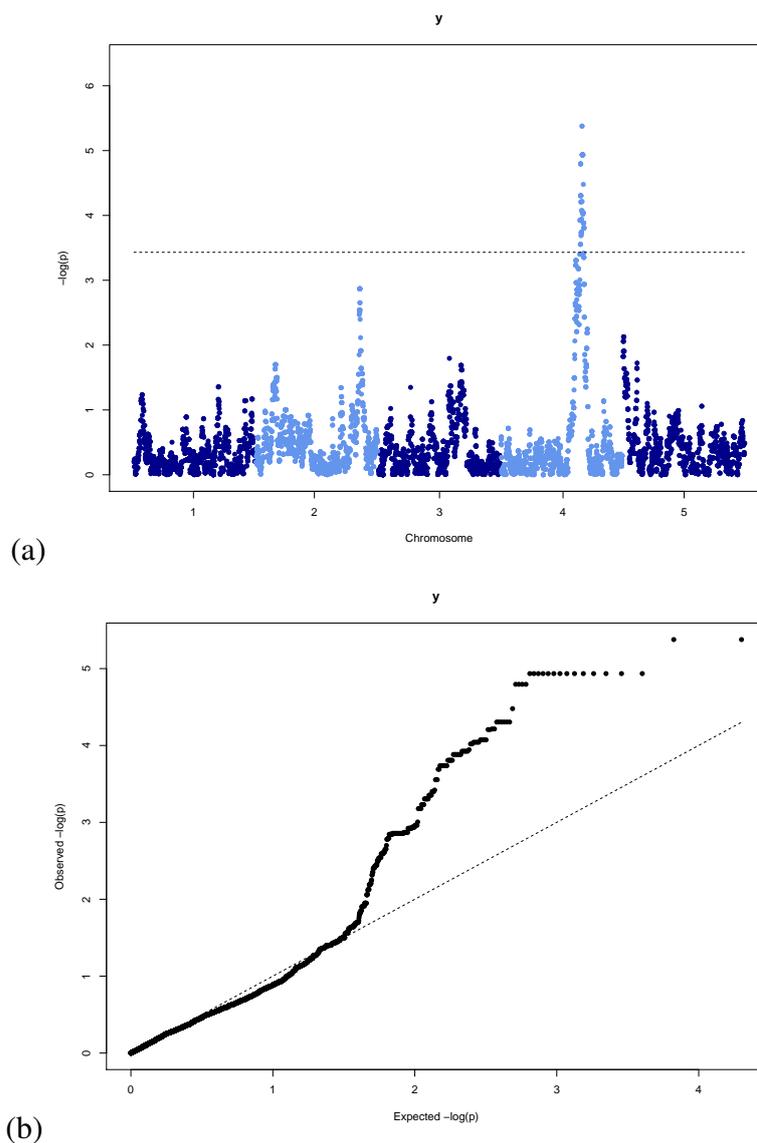
> head(myG)
marca  crom  pos G1 G2 G3 G4 G5 G6 G7 G8  G9 G10 G11 G12 G13 G14 G15 G16
1_1    1    0.0 0 -1  0  0  1  1  0  1  -1  0  0  0  0  0  0  -1
1_2    1    0.1 0 -1  0  0  1  1  0  1  -1  0  0  0  0  0  0  -1
1_3    1    0.3 0 -1  0  0  1  1  0  1  -1  0  0  0  0  0  0  -1

> dim(myG)
[1] 10010  303
```

Os nomes das colunas com as chamadas dos marcadores devem corresponder aos rótulos na primeira coluna do arquivo fenótipo. Carregar o pacote rrBLUP e usar sua função para executar a função GWAS, executado com a aproximação P3D:

```
library(rrBLUP)
resposta = GWAS(myY, myG, P3D=TRUE)
```

Figura 3 – Gráfico Manhattan (a) e QQ-plot (b) pela função GWAS pacote rrBLUP



Fonte: Da autora (2019)

No Gráfico de Manhattan (a) da Figura 3, a linha tracejada mostra o valor-p correspondente a uma taxa de falsa descoberta (FDR) de 0,05 e o gráfico QQ-plot (b) mostra a expectativa sob a hipótese nula.

A função GWAS tem a capacidade de executar o EMMA original (KANG et al., 2008) (Kang et al. 2008) como o métodos mais rápido, mas aproximado EMMAx/P3D (ZHANG et al., 2010), conforme declarado pela terminologia $P3D=TRUE$ (parâmetros populacionais previamente determinados). A opção $P3D=TRUE$ é mais rápida, mas pode subestimar a significância em comparação com $P3D=FALSE$. Mais detalhes sobre função GWAS, podem ser encontradas ao digitar help (“GWAS”) como recurso de script pelo *software* R.

GWAS com pacote GAPIT e pacote mrMLM

Uma ferramenta que apresenta métodos estatísticos para estudos de associação e predição para dados genômicos é o GAPIT (*Genome Association e Prediction Integrated Tool*). O pacote GAPIT, para uso pelo *software* R, implementa métodos estatísticos avançados, manipula grandes conjuntos de dados, com mais de 10000 indivíduos e 1 milhão de SNPs (LIPKA et al., 2012). Passa a ser um pacote alternativo, cujas funções apresentam considerável desempenho nas análises de grandes dados.

Além disso, o pacote GAPIT é uma preferência para escrita de funções breves, com estrutura de dados simples. Pode-se interpretar, analisar, acessar e investigar informações sobre os dados. As formas de visualização são reforçadas por métodos implementados no GAPIT (produz gráficos abrangentes, testes estatísticos, tabelas de qualidade e uma variedade de saídas) e torna-se importante ao aproveitar o poder de avaliações estatísticas de predições dos resultados genômicos.

A seguir, destacam-se algumas estruturas e funcionalidades do pacote GAPIT para o *software* R, que conduz estudos de GWS e GWAS.

GAPIT é um pacote executado pelo *software* R e pode ser obtido por meio do *download* em <http://www.r-project.org> ou <http://www.rstudio.com>. O usuário pode fornecer dados genotípicos, fenotípicos, matrizes de parentesco, estrutura de população e armazená-los em HapMap (formato comumente usado para armazenar sequencias de SNPs, cromossomos e posição) ou em um *data.frame*. Maiores detalhes podem ser obtidos pelo código R reproduzível com propósito ilustrativo de simulação e análise de dados, em anexo.

Ao inserir os dados de fenótipos e de matriz genótipos simulados, os resultados serão salvos no diretório de trabalho em arquivos de saída, em formatos (.csv) ou (.pdf), incluem resumos tanto para GWAS, quanto para GWS.

Apresenta-se um programa, no qual as linhas de comandos usa os arquivos dos dados genotípicos (myGD) e fenotípicos (myY). O arquivo myGM contém as informações dos SNPs, com colunas sendo identidade, cromossomo e posição.

```
myGD <- read.table("GD.txt",head=TRUE)      # matriz de genótipos
myY  <- read.table("yo2.txt",head=TRUE)     # vetor de fenótipos
myGM <- read.table("infsnp.txt",head=TRUE)  # SNPs, cromossomo, posição
```

O segundo passo é executar a função GAPIT e assim, os arquivos com resultados e gráficos serão salvos em um diretório.

```
myGAPIT <- GAPIT(
Y = myY,
GD = myGD,
GM = myGM,
group.from = nrow(myY),
group.to   = nrow(myY),
PCA.total  = 3,
Model.selection = TRUE)
```

O cálculo da matriz de parentesco pode ser feito por métodos como o de VanRaden (VANRADEN, 2008). O pacote GAPIT também oferece opção de calcular componentes principais, a partir dos dados genotípicos. As saídas são exibidas como arquivos que incluem resumos com resultados em GWAS e em GWS. Os resultados da GWAS são resumidos por gráficos de Manhattan, gráficos Q-Q plot e tabelas. Da mesma forma, os resultados da GWS são apresentados em um *Heatmap* e tabelas. São incluídos gráficos das estimativas de herdabilidade, da função de verossimilhança, dentro outros.

O número de componentes principais (*Principal Components-PCs*), incluídos nos modelos para GWAS, pode ser ajustado no GAPIT. Para determinar o número de PCs que explicam a estrutura da população, é importante determiná-las pela saída fornecida pelo GAPIT (se pelo menos um PC for selecionado para inclusão no modelo final). Indica-se a execução de seleção do modelo, baseado no Critério de Informação Bayesiano BIC (fornecido pela função GAPIT `Model.selection=TRUE`) (LIPKA et al., 2012).

Para facilitar o uso do pacote GAPIT, exemplo de dados, resultados, manual e código-fonte estão disponíveis em:<http://zzlab.net/GAPIT>.

As análises GWAS e GWS pelo GAPIT são realizadas recorrendo à implementação de métodos que incluem CMLM (*Compressed Mixed Linear Models*) (KANG et al., 2010), P3D ou EMMAx. O método P3D foi desenvolvido de forma independente por Kang et al. (2010) e é implementado em um programa C chamado EMMA *expedited* (EMMAx).

O método CMLM é um método existente para GWAS, de varredura de todo genoma, de um único locus. O CMLM diminui o tamanho efetivo da amostra ao agrupar indivíduos e elimina a necessidade de recalculando componentes de variância. Outros métodos como ECMLM, CMLM e SUPER foram implementados no pacote GAPIT e algumas das seis comparações metodológicas revisadas e avaliadas por Wen et al. (2017), seguem na tabela 56.

Tabela 56 – Comparação de alguns métodos para GWAS

| Caso | FASTmrEMMA | CMLM | ECMLM | SUPER |
|--|--|--|--|--|
| Modelo | multiloci | único locus | único locus | único locus |
| Efeito QTN | aleatório | fixo | fixo | fixo |
| Controle do <i>background</i> poligênico | sim | sim | sim | sim |
| Controle da estrutura populacional | sim | sim | sim | sim |
| Número de componentes de variância | três | dois | dois | dois |
| Variância poligênica residual | fixo | fixo | fixo | - |
| valor crítico | LOD=3 | valor-p=0,05 | valor-p=0,05/k | valor-p=0,05/k |
| Transformação da matriz | Matriz de covariância da matriz poligênica K é um ruído branco ambiental. Número de autovalores não nulos é especificado como um. | O parentesco entre os indivíduos é substituído pelo parentesco entre grupos. Ajusta os grupos como o efeito aleatório, estima os parâmetros populacionais apenas uma vez e depois os fixa para testar os marcadores genéticos. | Parentesco entre os indivíduos é substituído pelo parentesco entre os grupos. Escolhe-se a melhor combinação entre algoritmos de parentesco e algoritmos de agrupamento. | Reduz drasticamente o número de marcadores usados para definir relacionamentos individuais e os utiliza no FaST-LMM. |

QTN (*quantitative trait nucleotides*); LOD (logaritmo da probabilidade); *k* é o número de marcadores

Fonte: Adaptado (WEN et al., 2017)

Um método recente descrito por Wen et al. (2017) é o FASTmrEMMA (*fast multi-locus random-SNP-effect EMMA*), como uma abordagem em dois estágios para GWAS e fornece uma alternativa multiloci. O método foi implementado em R e pode ser baixado em <https://cran.r-project.org/web/packages/mrMLM/index.html>.

Com o pacote mrMLM, seis métodos para estudos de associação de genômica multiloci são implementados: mrMLM, FASTmrMLM, FASTmrEMMA, pLARmEB, pKWmEB e ISIS EM-BLASSO.