



ANDREZZA KÉLLEN ALVES PAMPLONA

**EFICIÊNCIA DE UM NOVO MÉTODO DE
IDENTIFICAÇÃO DE QTLs SOB ALTOS NÍVEIS DE
PERDAS DE MARCADORES**

LAVRAS - MG

2014

ANDREZZA KÉLLEN ALVES PAMPLONA

**EFICIÊNCIA DE UM NOVO MÉTODO DE IDENTIFICAÇÃO DE QTLs
SOB ALTOS NÍVEIS DE PERDAS DE MARCADORES**

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

Orientador

Dr. Júlio Sílvio de Sousa Bueno Filho

Coorientador

Dr. Marcio Balestre

**LAVRAS - MG
2014**

**Ficha Catalográfica Elaborada pela Coordenadoria de Produtos e
Serviços da Biblioteca Universitária da UFLA**

Pamplona, Andrezza Kéllen Alves.

Eficiência de um novo método de identificação de QTLs sob
altos níveis de perdas de marcadores / Andrezza Kéllen Alves
Pamplona. – Lavras : UFLA, 2014.

112 p. : il.

Dissertação (mestrado) - Universidade Federal de Lavras, 2014.

Orientador: Júlio Sílvio de Sousa Bueno Filho.

Bibliografia.

1. Regressão bayesiana. 2. Análise de QTL. 3. Múltiplas marcas.
4. Genome wide. I. Universidade Federal de Lavras. II. Título.

CDD - 519.542

ANDREZZA KÉLLEN ALVES PAMPLONA

**EFICIÊNCIA DE UM NOVO MÉTODO DE IDENTIFICAÇÃO DE QTLs
SOB ALTOS NÍVEIS DE PERDAS DE MARCADORES**

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

APROVADA em 27 de fevereiro de 2014.

Dr. João Bosco dos Santos
Dr. Marcio Balestre
Dra. Maria Imaculada de Sousa Silva

UFLA
UFLA
UFU

Dr. Júlio Sílvio de Sousa Bueno Filho
Orientador

**LAVRAS - MG
2014**

*Ao meu pai, Waldemar Pamplona da Silva, meu mestre, meu ídolo e minha
referência.*

*À minha mãe, Rozâna Alves da Silva Pamplona, minha força, meu apoio e
minha vida.*

À minha irmã, Greicy Kelly Alves Pamplona, minha companheira.

*Ao meu irmão (In Memoriam), Christian Darwin Alves Pamplona, que
mesmo ausente fisicamente, acredito estar sempre comigo espiritualmente.*

DEDICO

AGRADECIMENTOS

A Deus, o maior mestre, que permitiu que tudo isso acontecesse, me dando saúde e força para superar as dificuldades e me enviando boas vibrações.

Aos meus pais, *Waldemar Pamplona da Silva* e *Rozâna Alves da Silva Pamplona*, e à minha irmã, *Greicy Kelly Alves Pamplona*, pelo amor, incentivo e apoio incondicionais, por estarem sempre a postos quando preciso e por todo o esforço que me permitiu estar aqui.

Aos meus avós, *Waldiner Alves da Silva* e *Maria Emília Ramos da Silva*, que, mesmo distantes, me acompanharam nesta jornada com orações e pensamentos positivos, me alimentando de certezas, força e paciência.

Aos meus familiares, pelo carinho e preocupação, pelas orações e abraços e por compreenderem minha ausência em diversos momentos.

Ao meu namorado, *Carlos Pereira da Silva*, pela ajuda, paciência, apoio e compreensão dos momentos de ausência e nervosismo.

Aos meus amigos de longe, em especial *Weila Freitas* e *Rogério Reis dos Anjos*, pela preocupação demonstrada por meio de ligações, orações e emails.

Aos meus amigos e colegas de mestrado, em especial *Luciano Antonio de Oliveira*, *Fernando Ribeiro Cassiano* e *Joel Jorge Nuvunga*, pela amizade, apoio e ajuda em todos os momentos.

Ao meu orientador, *Júlio Sílvio de Sousa Bueno Filho*, por me orientar e proporcionar diversos conhecimentos.

Ao meu coorientador, *Marcio Balestre*, por se dispor a trabalhar comigo e pelo enorme auxílio neste trabalho no pouco tempo que lhe coube.

Aos membros da banca, *João Bosco dos Santos* e *Maria Imaculada de Sousa Silva*, pela disponibilidade e contribuições oferecidas neste trabalho.

Aos diversos *professores* do Departamento de Ciências Exatas da UFLA, que fizeram parte da minha formação acadêmica.

Aos funcionários do Departamento de Ciências Exatas, em especial à secretária de Pós-Graduação *Josiane Oliveira*, pela ajuda neste processo.

À Universidade Federal de Lavras (UFLA) e ao Departamento de Ciências Exatas (DEX), pela oportunidade de realização do mestrado.

Ao CNPQ - Conselho Nacional de Desenvolvimento Científico e Tecnológico - pela concessão da bolsa de estudos.

A todos que, diretamente ou indiretamente, me apoiaram e ajudaram nesta jornada.

*"Sonhos determinam o que você quer.
Ação determina o que você conquista."
(Aldo Novak)*

RESUMO

Em diversas espécies, o baixo nível de polimorfismo impede a construção de mapas de ligação que possam ser usados na identificação de QTLs no genoma. Objetivou-se neste trabalho comparar dois métodos de identificação de QTLs que não requerem mapas de ligação em estudos de associação. O Método I é o da regressão bayesiana de múltiplos marcadores, originalmente proposto por Xu (2003). O Método II consiste em uma adaptação do Método I e do método de Wang et al. (2005), porém utilizando o conceito descrito por Doerge, Zeng e Weir (1997). Nesse método, os marcadores não são regredidos diretamente sobre o fenótipo, mas servem como pivôs para a busca do QTL ao longo do genoma - se tem, então, um mapeamento de múltiplos QTLs. Para verificar a efetividade do método, realizou-se simulação de 300 indivíduos pertencentes à população F_2 , com dois níveis de perdas de marcadores (20% e 80%), em um total de 165 marcadores, distribuídos em 11 cromossomos. Ao longo desses cromossomos, sete QTLs foram simulados. Foi analisado, também, um exemplo com dados reais envolvendo 186 progênies $F_{2:4}$ de feijão, com 59 marcadores, sendo 17 SSRs, 31 AFLPs e 11 SRAPs. No estudo de simulação, o Método II foi melhor que o Método I em ambos os níveis de perda de marcadores. Nos dados reais, o Método II detectou 17 marcadores promissores enquanto o Método I não detectou nenhum. O Método II mostrou maior poder de detecção e pode ser recomendado para estudos posteriores com dados reais e com outros delineamentos de cruzamento.

Palavras-chave: Regressão bayesiana. Análise de QTL. Múltiplas marcas. Genome wide.

ABSTRACT

In several species, a low level of polymorphism prevents the construction of linkage maps that can be used in the identification of QTLs in the genome. The objective of this study is to compare two methods of identifying QTLs that don't require genetic maps in association studies. The Method I is the multiple-markers Bayesian regression, originally proposed by Xu (2003). The Method II consists of an adaptation of the Method I and method of Wang et al. (2005), but using the concept described by Doerge, Zeng and Weir (1997). In this method, the markers are not directly regressed on phenotype, but serve as pivots for the search of the QTL along the genome - then it has multiple QTL mapping. To verify the effectiveness of the method, a simulation was carried out with 300 individuals belonging to the F_2 population in two levels of markers loss (20% and 80%) from a total of 165 markers, divided into 11 chromosomes. Throughout these chromosomes, seven QTLs were simulated. It was also considered an example with real data involving 186 $F_{2:4}$ progenies of beans with 59 markers, 17 SSRs, 31 AFLPs and 11 SRAPs. In the simulation's study, the Method II was better than the Method I in both levels of markers loss. In real data, the Method II detected 17 promising markers while the Method I didn't detect any. The Method II showed greater power to detect and it can be recommended for further studies with actual data and other designs crossover and genome wide.

Keywords: Bayesian regression. Analysis of QTL. Multiple markers. Genome wide.

LISTA DE FIGURAS

| | | |
|-----------|--|----|
| Figura 1 | Mapa genético simulado com onze grupos de ligação, 165 marcadores e sete QTLs. | 32 |
| Figura 2 | Poder dos Métodos I e II para detectar QTL, para o primeiro grupo de ligação simulado, com 20% de perda de marcadores. | 48 |
| Figura 3 | Poder dos Métodos I e II para determinar efeito aditivo dos QTLs, para o primeiro grupo de ligação, com 20% de perda de marcadores. | 49 |
| Figura 4 | QTLs estimados pelos Métodos I e II com as respectivas posições (em unidade Morgan) no cromossomo um, frequências relativas com intervalos de confiança exato para proporção com 5% de significância e diferenças absolutas dos efeitos aditivos entre os QTLs estimados e os QTLs simulados, para perda de 20% das marcas. | 51 |
| Figura 5 | Poder dos Métodos I e II para detectar QTL, para o quinto grupo de ligação simulado, com 20% de perda de marcadores. | 52 |
| Figura 6 | Poder dos Métodos I e II para determinar efeito aditivo dos QTLs, para o quinto grupo de ligação, com 20% de perda de marcadores. | 53 |
| Figura 7 | QTLs estimados pelos Métodos I e II com as respectivas posições (em unidade Morgan) no cromossomo cinco, frequências relativas com intervalos de confiança exato para proporção com 5% de significância e diferenças absolutas dos efeitos aditivos entre os QTLs estimados e os QTLs simulados, para perda de 20% das marcas. | 54 |
| Figura 8 | Poder dos Métodos I e II para detectar QTL, para o sétimo grupo de ligação simulado, com 20% de perda de marcadores. | 56 |
| Figura 9 | Poder dos Métodos I e II para determinar efeito aditivo dos QTLs, para o sétimo grupo de ligação, com 20% de perda de marcadores. | 57 |
| Figura 10 | QTLs estimados pelos Métodos I e II com as respectivas posições (em unidade Morgan) no cromossomo sete, frequências relativas com intervalos de confiança exato para proporção com 5% de significância e diferenças absolutas dos efeitos aditivos entre os QTLs estimados e os QTLs simulados, para perda de 20% das marcas. | 58 |
| Figura 11 | Poder dos Métodos I e II para detectar QTL, para o oitavo grupo de ligação simulado, com perda de 20% de marcadores. | 60 |

| | | |
|-----------|---|----|
| Figura 12 | Poder dos Métodos I e II para detectar QTL, para o oitavo grupo de ligação simulado, com perda de 20% de marcadores. | 61 |
| Figura 13 | QTLs estimados pelos Métodos I e II com as respectivas posições (em unidade Morgan) no cromossomo oito, frequências relativas com intervalos de confiança exato para proporção com 5% de significância e diferenças absolutas dos efeitos aditivos entre os QTLs estimados e os QTLs simulados, para perda de 20% das marcas. | 62 |
| Figura 14 | Poder dos Métodos I e II para detectar QTL, para o nono grupo de ligação simulado, com 20% de perda de marcadores. | 63 |
| Figura 15 | Poder dos Métodos I e II para determinar efeito aditivo dos QTLs, para o nono grupo de ligação, com 20% de perda de marcadores. | 64 |
| Figura 16 | QTLs estimados pelos Métodos I e II com as respectivas posições (em unidade Morgan) no cromossomo nove, frequências relativas com intervalos de confiança exato para proporção com 5% de significância e diferenças absolutas dos efeitos aditivos entre os QTLs estimados e os QTLs simulados, para perda de 20% das marcas. | 65 |
| Figura 17 | Poder dos Métodos I e II para detectar QTL, para o primeiro grupo de ligação simulado, com 80% de perda de marcadores. | 67 |
| Figura 18 | Poder dos Métodos I e II para determinar efeito aditivo dos QTLs, para o primeiro grupo de ligação, com 80% de perda de marcadores. | 68 |
| Figura 19 | QTLs estimados pelos Métodos I e II com as respectivas posições (em unidade Morgan) no cromossomo um, frequências relativas com intervalos de confiança exato para proporção com 5% de significância e diferenças absolutas dos efeitos aditivos entre os QTLs estimados e os QTLs simulados, para perda de 80% das marcas. | 69 |
| Figura 20 | Poder dos Métodos I e II para detectar QTL, para o quinto grupo de ligação simulado, com 80% de perda de marcadores. | 70 |
| Figura 21 | Poder dos Métodos I e II para determinar efeito aditivo dos QTLs, para o quinto grupo de ligação, com 80% de perda de marcadores. | 71 |

| | | |
|-----------|--|----|
| Figura 22 | QTLs estimados pelos Métodos I e II com as respectivas posições (em unidade Morgan) no cromossomo cinco, frequências relativas com intervalos de confiança exato para proporção com 5% de significância e diferenças absolutas dos efeitos aditivos entre os QTLs estimados e os QTLs simulados, para perda de 80% das marcas. | 72 |
| Figura 23 | Poder dos Métodos I e II para detectar QTL, para o sétimo grupo de ligação simulado, com 80% de perda de marcadores. | 73 |
| Figura 24 | Poder dos Métodos I e II para determinar efeito aditivo dos QTLs, para o sétimo grupo de ligação, com 80% de perda de marcadores. | 74 |
| Figura 25 | QTLs estimados pelos Métodos I e II com as respectivas posições (em unidade Morgan) no cromossomo sete, frequências relativas com intervalos de confiança exato para proporção com 5% de significância e diferenças absolutas dos efeitos aditivos entre os QTLs estimados e os QTLs simulados, para perda de 80% das marcas. | 75 |
| Figura 26 | Poder dos Métodos I e II para detectar QTL, para o oitavo grupo de ligação simulado, com 80% de perda de marcadores. | 76 |
| Figura 27 | Poder dos Métodos I e II para determinar efeito aditivo dos QTLs, para o oitavo grupo de ligação, com 80% de perda de marcadores. | 77 |
| Figura 28 | QTLs estimados pelos Métodos I e II com as respectivas posições (em unidade Morgan) no cromossomo oito, frequências relativas com intervalos de confiança exato para proporção com 5% de significância e diferenças absolutas dos efeitos aditivos entre os QTLs estimados e os QTLs simulados, para perda de 80% das marcas. | 78 |
| Figura 29 | Poder dos Métodos I e II para detectar QTL, para o nono grupo de ligação simulado, com 80% de perda de marcadores. | 79 |
| Figura 30 | Poder dos Métodos I e II para determinar efeito aditivo dos QTLs, para o nono grupo de ligação, com 80% de perda de marcadores. | 80 |
| Figura 31 | QTLs estimados pelos Métodos I e II com as respectivas posições (em unidade Morgan) no cromossomo nove, frequências relativas com intervalos de confiança exato para proporção com 5% de significância e diferenças absolutas dos efeitos aditivos entre os QTLs estimados e os QTLs simulados, para perda de 80% das marcas. | 81 |

| | | |
|-----------|--|----|
| Figura 32 | Poder dos Métodos I e II para detectar QTL, para o nono grupo de ligação simulado, com 80% de perda de marcadores. | 82 |
| Figura 33 | Poder dos Métodos I e II para detectar QTL, para o nono grupo de ligação simulado, com 80% de perda de marcadores. | 82 |
| Figura 34 | Poder dos Métodos I e II para determinar efeito dominante dos QTLs, para o primeiro grupo de ligação simulado, com 20% de perda de marcadores. | 90 |
| Figura 35 | Poder dos Métodos I e II para determinar efeito dominante dos QTLs, para o quinto grupo de ligação simulado, com 20% de perda de marcadores. | 91 |
| Figura 36 | Poder dos Métodos I e II para determinar efeito dominante dos QTLs, para o sétimo grupo de ligação simulado, com 20% de perda de marcadores. | 91 |
| Figura 37 | Poder dos Métodos I e II para determinar efeito dominante dos QTLs, para o oitavo grupo de ligação simulado, com 20% de perda de marcadores. | 92 |
| Figura 38 | Poder dos Métodos I e II para determinar efeito dominante dos QTLs, para o nono grupo de ligação simulado, com 20% de perda de marcadores. | 92 |
| Figura 39 | Poder dos Métodos I e II para determinar efeito dominante dos QTLs, para o primeiro grupo de ligação simulado, com 80% de perda de marcadores. | 93 |
| Figura 40 | Poder dos Métodos I e II para determinar efeito dominante dos QTLs, para o quinto grupo de ligação simulado, com 80% de perda de marcadores. | 93 |
| Figura 41 | Poder dos Métodos I e II para determinar efeito dominante dos QTLs, para o sétimo grupo de ligação simulado, com 80% de perda de marcadores. | 94 |
| Figura 42 | Poder dos Métodos I e II para determinar efeito dominante dos QTLs, para o oitavo grupo de ligação simulado, com 80% de perda de marcadores. | 94 |
| Figura 43 | Poder dos Métodos I e II para determinar efeito dominante dos QTLs, para o nono grupo de ligação simulado, com 80% de perda de marcadores. | 95 |

| | | |
|-----------|--|-----|
| Figura 44 | QTLs estimados pelos Métodos I e II com as respectivas posições (em unidade Morgan) no cromossomo um, frequências relativas com intervalos de confiança exato para proporção com 5% de significância e diferenças absolutas dos efeitos dominantes entre os QTLs estimados e os QTLs simulados, para perda de 20% das marcas. | 97 |
| Figura 45 | QTLs estimados pelos Métodos I e II com as respectivas posições (em unidade Morgan) no cromossomo cinco, frequências relativas com intervalos de confiança exato para proporção com 5% de significância e diferenças absolutas dos efeitos dominantes entre os QTLs estimados e os QTLs simulados, para perda de 20% das marcas. | 98 |
| Figura 46 | QTLs estimados pelos Métodos I e II com as respectivas posições (em unidade Morgan) no cromossomo sete, frequências relativas com intervalos de confiança exato para proporção com 5% de significância e diferenças absolutas dos efeitos dominantes entre os QTLs estimados e os QTLs simulados, para perda de 20% das marcas. | 99 |
| Figura 47 | QTLs estimados pelos Métodos I e II com as respectivas posições (em unidade Morgan) no cromossomo oito, frequências relativas com intervalos de confiança exato para proporção com 5% de significância e diferenças absolutas dos efeitos dominantes entre os QTLs estimados e os QTLs simulados, para perda de 20% das marcas. | 100 |
| Figura 48 | QTLs estimados pelos Métodos I e II com as respectivas posições (em unidade Morgan) no cromossomo nove, frequências relativas com intervalos de confiança exato para proporção com 5% de significância e diferenças absolutas dos efeitos dominantes entre os QTLs estimados e os QTLs simulados, para perda de 20% das marcas. | 101 |
| Figura 49 | QTLs estimados pelos Métodos I e II com as respectivas posições (em unidade Morgan) no cromossomo um, frequências relativas com intervalos de confiança exato para proporção com 5% de significância e diferenças absolutas dos efeitos dominantes entre os QTLs estimados e os QTLs simulados, para perda de 80% das marcas. | 102 |

| | | |
|-----------|--|-----|
| Figura 50 | QTLs estimados pelos Métodos I e II com as respectivas posições (em unidade Morgan) no cromossomo cinco, frequências relativas com intervalos de confiança exato para proporção com 5% de significância e diferenças absolutas dos efeitos dominantes entre os QTLs estimados e os QTLs simulados, para perda de 80% das marcas. | 103 |
| Figura 51 | QTLs estimados pelos Métodos I e II com as respectivas posições (em unidade Morgan) no cromossomo sete, frequências relativas com intervalos de confiança exato para proporção com 5% de significância e diferenças absolutas dos efeitos dominantes entre os QTLs estimados e os QTLs simulados, para perda de 80% das marcas. | 104 |
| Figura 52 | QTLs estimados pelos Métodos I e II com as respectivas posições (em unidade Morgan) no cromossomo oito, frequências relativas com intervalos de confiança exato para proporção com 5% de significância e diferenças absolutas dos efeitos dominantes entre os QTLs estimados e os QTLs simulados, para perda de 80% das marcas. | 105 |
| Figura 53 | QTLs estimados pelos Métodos I e II com as respectivas posições (em unidade Morgan) no cromossomo nove, frequências relativas com intervalos de confiança exato para proporção com 5% de significância e diferenças absolutas dos efeitos dominantes entre os QTLs estimados e os QTLs simulados, para perda de 80% das marcas. | 106 |

LISTA DE TABELAS

| | | |
|----------|--|-----|
| Tabela 1 | Posições, efeitos aditivos e efeitos dominantes dos QTLs simulados nos respectivos grupos de ligação. | 31 |
| Tabela 2 | Frequências de detecção, total de seleções e médias das distâncias para os dois métodos, por grupo de ligação (GL), com perda de 20% das marcas. | 107 |
| Tabela 3 | Frequências de detecção, total de seleções e médias das distâncias para os dois métodos, por grupo de ligação (GL), com perda de 80% das marcas. | 110 |

SUMÁRIO

| | | |
|--------------|---|----|
| 1 | INTRODUÇÃO | 18 |
| 2 | REFERENCIAL TEÓRICO | 21 |
| 2.1 | Mapeamento de QTLs | 21 |
| 2.2 | Método <i>moving away from the markers</i> | 26 |
| 2.3 | Abordagem bayesiana <i>shrinkage</i> | 29 |
| 3 | MATERIAL E MÉTODOS | 31 |
| 3.1 | Simulação | 31 |
| 3.2 | Mapeamento por múltiplos marcadores | 33 |
| 3.2.1 | Modelo Linear | 33 |
| 3.2.2 | Estimação Bayesiana | 34 |
| 3.2.3 | Amostragem das distribuições condicionais completas | 35 |
| 3.3 | <i>Moving away from the marker</i> com análise bayesiana | 37 |
| 3.3.1 | Modelo Linear | 38 |
| 3.3.2 | Estimação Bayesiana | 38 |
| 3.3.3 | Amostragem das distribuições condicionais completas | 40 |
| 3.4 | Análise pós-MCMC | 44 |
| 3.5 | Dados do experimento real | 45 |
| 3.6 | Análise dos resultados | 45 |
| 4 | RESULTADOS | 47 |
| 4.1 | Dados simulados | 47 |
| 4.1.1 | Perda de 20% das marcas | 47 |
| 4.1.2 | Perda de 80% das marcas | 66 |
| 4.2 | Dados reais | 82 |
| 5 | DISCUSSÃO | 83 |
| 6 | CONCLUSÃO | 86 |
| | REFERÊNCIAS | 87 |
| | APÊNDICES | 90 |

1 INTRODUÇÃO

O estudo genético detalhado da expressão fenotípica de caracteres quantitativos é o grande interesse de geneticistas e melhoristas de diversas áreas. Supõe-se que a variação genética desses caracteres seja controlada pela segregação simultânea de muitos genes, distribuídos ao longo do genoma, em regiões definidas como QTLs (*Quantitative Trait Loci*), responsáveis pela expressão fenotípica.

Com grande quantidade de marcadores moleculares é possível construir mapas genéticos que, mediante alguns procedimentos estatísticos, permitam realizar o mapeamento de QTLs, ou seja, localizá-los em grupos de ligação, quantificar seus efeitos, determinar o número de QTLs envolvidos e sua distribuição no genoma. Segundo Moraes (2005, p. 9), "este conjunto de informações genéticas é um poderoso recurso no estudo da herança desses caracteres, criando novas perspectivas para aperfeiçoar os métodos de seleção e melhoramento".

Existem diversos métodos de mapeamento de QTLs, os quais diferem em requisitos computacionais, eficiência em termos de informação extraída, flexibilidade no que diz respeito a lidar com diferentes estruturas de dados e capacidade de mapear múltiplos QTLs.

O Mapeamento de Marcas Simples, proposto por Edwards, Stuber e Wendel (1987), utilizando análise de mínimos quadrados ou teste t , é eficiente em termos de velocidade computacional, mas não pode extrair todas as informações com base nos dados. O método de Mapeamento por Intervalo, baseado na análise de máxima verossimilhança, proposto por Lander e Botstein (1989), é um dos mais utilizados para análise de QTLs em cruzamentos controlados ou pedigrees estruturados. Este método foi estendido para Mapeamento por Intervalo Composto (ZENG, 1994) e Mapeamento por Múltiplos Intervalos (KAO; ZENG; TEASDALE, 1999). Outros métodos de mapeamento de QTL foram propostos: Ma-

peamento por Múltiplas Marcas (XU, 2003) e Mapeamento por Múltiplos QTLs (WANG et al., 2005).

Outra técnica de mapeamento de QTL, com marcas simples, foi proposta por Doerge, Zeng e Weir (1997). Esta técnica foi, posteriormente, denominada por Wu, Ma e Casella (2007) como *Moving Away From The Marker* e é baseada no teste de razão de verossimilhanças, considerando a hipótese de que o marcador e o QTL não estão ligados. Esta técnica pode ser estendida adotando-se a análise bayesiana e adaptando o mapeamento por múltiplos QTLs de Wang et al. (2005) juntamente com o mapeamento por múltiplas marcas de Xu (2003).

A abordagem bayesiana, para mapeamento de QTLs, tornou-se uma ferramenta poderosa em estudos de análise genômica. Permite grande flexibilidade de análise em modelos complexos e, também, em situações onde o número de parâmetros a ser estimado é maior que o número de observações, sendo possível incorporar a incerteza sobre as quantidades desconhecidas, por exemplo, com relação ao número de QTLs (XU, 2003).

Alguns dos métodos citados acima necessitam de um genoma saturado, ou seja, alta quantidade de marcadores distribuídos em todo o genoma. Porém, o baixo nível de polimorfismo, em diversas espécies, como a do feijão, torna o genoma pobremente saturado, o que impede a construção de mapas de ligação que possam ser usados na identificação de QTLs no genoma. Por isso, é importante saber qual nível de saturação do genoma é considerado aceitável para garantir o poder dos métodos de mapeamento.

O presente trabalho foi realizado com os seguintes objetivos:

- 1 Propor um método de análise para a técnica *moving away from the marker*, utilizando abordagem bayesiana.
- 2 Comparar as versões bayesianas do mapeamento por regressão linear de

múltiplas marcas e do mapeamento *moving away from the marker* proposto, sob diferentes níveis de perda de marcadores.

- 3 Aplicar os métodos em dados reais para validar os resultados encontrados.

2 REFERENCIAL TEÓRICO

Quantitative trait loci são regiões do genoma responsáveis pela expressão de caracteres fenotípicos. Com o advento dos marcadores moleculares tornou-se possível mapear regiões dos cromossomos que afetam tais caracteres quantitativos. Vários são os métodos disponíveis na literatura para mapear os QTLs, sendo alguns deles discutidos nesta seção.

2.1 Mapeamento de QTLs

Mapear QTLs significa identificá-los e localizá-los no genoma, estimando suas posições e efeitos genéticos (efeito aditivo, efeito de dominância, epistasia, pleiotropia e outros efeitos que estiverem presentes no modelo). É necessário obter uma coleta acurada de dados do caráter quantitativo de interesse e dos marcadores moleculares dentro da população experimental (CAMP; COX, 2002). Segundo Toledo et al. (2008), após obter os dados citados acima, deve-se realizar três passos para o mapa de ligação:

- i Distribuir os marcadores utilizados em grupos de ligação.
- ii Ordená-los no grupo.
- iii Dentro de cada grupo, estimar a distância entre os marcadores adjacentes ou flaqueadores.

Uma vez construído o mapa de ligação, pode-se dar continuidade ao mapeamento de QTLs, utilizando métodos disponíveis na literatura para correlacionar os dados fenotípicos com os dados genotípicos.

O mapeamento por marcas simples, proposto por Edwards, Stuber e Wendel (1987), associa um marcador por vez ao caráter de interesse e consiste na

comparação estatística das médias dos fenótipos dentro de cada classe genotípica do marcador. A análise pode ser realizada, por meio de teste t , regressão linear simples e análise de variância, testando-se a hipótese nula de que as classes genotípicas não diferem em fenótipo para um determinado marcador molecular. Um resultado significativo sugere a possibilidade de, pelo menos um QTL, estar ligado à marca. Cada teste de associação carácter-marcador é realizado, independente da informação dos outros marcadores e, portanto, se um cromossomo possui n marcas, n testes deverão ser realizados (MORAES, 2005). Segundo Lander e Botstein (1989), algumas desvantagens deste mapeamento são: (i) os efeitos fenotípicos de QTLs são seriamente subestimados; (ii) o número de progênies requerido para detectar QTLs é maior que o necessário; (iii) a provável posição do QTL não é bem definida, não podendo distinguir entre ligação muito próxima com um QTL de pequeno efeito ou ligação mais distante com um QTL de alto efeito. No entanto, esta análise não requer mapa genético e é útil para um ensaio preliminar da existência de um QTL.

Lander e Botstein (1989) propuseram o mapeamento por intervalo simples, no qual se verifica a presença de QTL em um intervalo definido por marcas adjacentes. Assim, em um intervalo, várias posições são testadas para a presença do QTL por meio do teste da razão de verossimilhanças, que compara a função de verossimilhança da hipótese nula (ausência de QTL) com a hipótese alternativa (presença de QTL na posição sendo testada). Um QTL é dito estar presente, quando o *LOD* score (*log of odds score*) maximizado excede algum valor limiar pré-determinado (*threshold*). Em contraste com a abordagem anterior, as localizações e os efeitos estimados dos QTLs tendem a ser, assintoticamente, não viesados se existe somente um QTL segregando em um cromossomo. Este método requer menos progênies para a detecção de QTLs, possui maior poder de detecção e maior

precisão na estimativa dos parâmetros associados ao QTL (ZENG, 1994). Porém, se os marcadores estão muito distantes um do outro, esta análise é imprecisa na detecção de um QTL. Como cada QTL é ajustado individualmente no modelo, os outros possíveis QTLs presentes no genoma são ignorados, aumentando o resíduo. Além disso, eventuais QTLs que estejam localizados perto do intervalo que está sendo testado interferem no processo de estimação, levando a detectar erroneamente um QTL neste intervalo, quando na realidade não há nenhum presente, os chamados "QTLs fantasmas"(MARTINEZ; CURNOW, 1992).

Jansen (1993) e Zeng (1993, 1994), independentemente, propuseram a ideia de combinar o mapeamento por intervalo sobre duas marcas com análise de regressão múltipla sobre as demais marcas, visando eliminar o problema dos QTLs "fantasmas". Zeng (1994) nomeou esta combinação de mapeamento por intervalo composto, por se tratar de um enfoque misto entre as técnicas de regressão e o método da razão de verossimilhanças. Com isso, utiliza-se um intervalo formado por dois marcadores adjacentes para localizar a posição de teste e, ao mesmo tempo, utiliza as outras marcas como covariáveis na regressão de fenótipos em genótipos para controlar os outros QTLs e reduzir a variância residual (KAO; ZENG; TEASDALE, 1999), aumentando, assim, o poder do teste e diminuindo, consideravelmente, os QTLs "fantasmas", que só surgirão se houver QTLs nos intervalos adjacentes fortemente ligados ao intervalo sendo estudado (ZENG, 1994).

Um método de mapeamento mais complexo foi proposto por Kao, Zeng e Teasdale (1999), chamado mapeamento por múltiplos intervalos. Neste caso, intervalos múltiplos formados por dois marcadores adjacentes são usados, simultaneamente, para ajustar múltiplos prováveis QTLs, diretamente, no modelo. Este método é baseado no modelo de Cockerham (KAO; ZENG, resultados não publicados) (KAO; ZENG; TEASDALE, 1999) para interpretar os parâmetros genéti-

cos e no método da máxima verossimilhança para estimá-los. Um processo de seleção *stepwise* com estatística de teste de razão de verossimilhanças como um critério é proposto para identificar QTLs. Segundo os autores, este método tende a ser mais poderoso e preciso na detecção de QTLs. De acordo com Xu (2003), uma vez que este método não tem uma maneira conveniente de lidar com muitos intervalos ao mesmo tempo, os intervalos devem ser incluídos seletivamente no modelo, gerando um problema de seleção de modelo simultaneamente à identificação de QTLs. O critério de incluir ou excluir um QTL pode ser arbitrário e o espaço amostral dos possíveis modelos pode ser tão grande que dificilmente é totalmente explorado. Por este motivo, este método foi denominado por Xu (2003) de mapeamento seletivo de múltiplos intervalos e o verdadeiro mapeamento por múltiplos intervalos deveria incluir todos os intervalos definidos por marcadores simultaneamente.

Alguns dos métodos citados acima necessitam de mapas de ligação para identificação de QTLs. Tais mapas são mais bem construídos em genomas altamente saturados. Porém, quando o genoma é pobremente saturado de marcadores, a construção do mapa torna-se imprecisa e inviável, já que não se consegue reconstruir os grupos de ligação ou constroem-se mapas de ligação distorcidos que nem sempre representam o mapa físico. Poder-se-ia argumentar, nessa situação, que seria possível o uso de um mapa de consenso. Contudo, essa abordagem só é viável se os marcadores estiverem bem distribuídos ao longo dos grupos de ligação. Assim, torna-se necessária uma abordagem de busca de QTLs ao longo do genoma que não exija mapa. Uma alternativa seria o uso da regressão múltipla de marcadores apenas para análise de associação. Se o genoma é altamente saturado, a regressão é assintoticamente eficiente (XU, 2003); em outros casos, pode ocorrer viés.

Assim, um método de mapeamento que não utiliza mapa de ligação foi proposto por Xu (2003), chamado mapeamento por múltiplas marcas, que aplicou a abordagem bayesiana com o objetivo de avaliar simultaneamente os efeitos dos QTLs associados com os marcadores de todo o genoma, adotando um método de regressão bayesiano *shrinkage*, sendo capaz de manusear situações em que o número de parâmetros a ser estimado é maior que o número de observações. Nesta abordagem, cada marcador é assumido como um provável QTL, fazendo com que a matriz de incidência (determinada pelos genótipos dos marcadores) seja integralmente observável. A priori, cada QTL é uma variável aleatória proveniente de uma distribuição normal com média zero e variância individual. Dessa forma, cada possível QTL é penalizado pela razão da sua variância com a variância residual. Assim, aqueles QTLs de pequeno efeito e baixa variância têm seus efeitos "encolhidos" a valores próximos de zero; já aqueles de grande efeito tendem a apresentar variância genética alta e são menos penalizados (BALESTRE et al., 2012).

De acordo com Wang et al. (2005, p. 466), "mapear QTLs, entretanto, não é tão simples devido à incerteza dos genótipos dos QTLs e, conseqüentemente, a matriz de incidência não é mais observada. Mais ainda, as posições dos QTLs são parâmetros adicionais de interesse". Baseado nisso, estes autores propuseram estender a estimação bayesiana *shrinkage* de Xu (2003) para mapeamento de QTLs em que as posições e os efeitos dos QTLs são estimados simultaneamente. Este método assume que cada intervalo definido pelos marcadores adjacentes possui um QTL. Alguns destes QTLs têm grandes efeitos, mas a maioria tem efeito zero, ou seja, não existe QTL naquele intervalo. Em vez de excluir QTLs com efeitos zero do modelo, como feito em qualquer abordagem de seleção de variáveis, este método inclui estes QTLs no modelo. Segundo Wang et al. (2005, p. 478), "é importante incluir estes pequenos QTLs no modelo porque, coletivamente, podem

ter uma contribuição significativa para a variância da característica. Se eles são ignorados, a variância residual será inflacionada, o que não é desejável". A técnica que permite que os efeitos destes QTLs nulos sejam estimados próximos de zero é a abordagem de estimação *shrinkage* seletiva.

Para Wang *et al.* (2005, p. 479), seu método é "mais simples e mais fácil de programar que qualquer outro método de seleção de modelo". Porém, pode apresentar alguns problemas: não existe flexibilidade para incorporar conhecimento a priori sobre o número de QTLs e as posições dos QTLs; não existe uma afirmação probabilística explícita sobre quantas vezes um QTL é incluído no modelo, embora a densidade a posteriori de um efeito de QTL possa ser usada para extrair tal informação; e, o tempo computacional requerido para a estimação bayesiana *shrinkage* é, normalmente, longo por causa do grande número de efeitos de QTLs incluídos no modelo.

2.2 Método *moving away from the markers*

Doerge, Zeng e Weir (1997) fizeram uma revisão das principais contribuições na área de estatística aplicada ao uso da tecnologia de marcadores moleculares e genética quantitativa na busca de genes que afetam características quantitativas. Uma das análises ilustradas no referido artigo é a de marcador simples, porém não se assume que o marcador é um provável QTL. Esta técnica foi, posteriormente, denominada por Wu, Ma e Casella (2007) como *Moving Away From The Marker*. Nesse método, testa-se a hipótese de que o marcador não está ligado com o provável QTL (a frequência de recombinação entre ambos é 0,5). A rejeição desta hipótese sugere que o caráter é afetado por um QTL próximo ao marcador (DOERGE; ZENG; WEIR, 1997).

Ilustrando a técnica com um delineamento de retrocruzamento em que

existem dois genótipos em cada marcador ou QTL, o genótipo observado será ou M_1M_1 ou M_1M_2 , mas, dados estes genótipos para os marcadores, o genótipo do QTL será ou Q_1Q_1 ou Q_1Q_2 com probabilidades dadas abaixo, sendo r a fração de recombinação entre o marcador e o QTL:

$$\begin{aligned} P(M_1Q_1/M_1Q_1) &= \frac{1-r}{2} & P(M_1Q_1/M_1Q_2) &= \frac{r}{2} \\ P(M_1Q_1/\{M_2Q_1\}) &= \frac{r}{2} & P(M_1Q_1/M_2Q_2) &= \frac{1-r}{2} \end{aligned}$$

Assumindo que o caráter fenotípico (y) segue uma distribuição normal, as distribuições do caráter esperado dentro destas duas classes são:

$$\begin{aligned} M_1/M_1 &: (1-r)N(\mu_1, \sigma^2) + rN(\mu_2, \sigma^2) \\ M_1/M_2 &: rN(\mu_1, \sigma^2) + (1-r)N(\mu_2, \sigma^2) \end{aligned}$$

sendo μ_1 e μ_2 as médias fenotípicas do caráter para genótipos Q_1Q_1 e Q_1Q_2 , respectivamente.

As médias e variâncias destas duas distribuições de mistura são:

$$\begin{aligned} \mu_{M_1/M_1} &= (1-r)\mu_1 + r\mu_2 \\ \mu_{M_1/M_2} &= r\mu_1 + (1-r)\mu_2 \\ \sigma_{M_1/M_1}^2 &= \sigma_{M_1/M_2}^2 = \sigma^2 + r(1-r)(\mu_1 - \mu_2)^2 \end{aligned}$$

Se não existe ligação entre o marcador e o QTL, ou seja, se $r = \frac{1}{2}$, então μ_{M_1/M_1} e μ_{M_1/M_2} são iguais. Assim, a hipótese nula de não ligação é:

$$H_0 : r = \frac{1}{2} \text{ ou } H_0 : \mu_{M_1/M_1} - \mu_{M_1/M_2} = (1-2r)(\mu_1 - \mu_2) = 0$$

É importante notar que sob H_0 não se pode dizer se $r = \frac{1}{2}$ ou se $\mu_1 = \mu_2$.

Se se assume que y_1, \dots, y_{n_1} são do grupo de marcadores M_1M_1 e que

y_{n_1+1}, \dots, y_n são do grupo de marcadores M_1M_2 , então a função de verossimilhança baseada no modelo de mistura é:

$$L(\mu_1, \mu_2, \sigma^2, r|y) = \prod_{i=1}^{n_1} (1-r) f(y_i|\mu_1, \sigma^2) + r f(y_i|\mu_2, \sigma^2) \\ \times \prod_{i=n_1+1}^n r f(y_i|\mu_1, \sigma^2) + (1-r) f(y_i|\mu_2, \sigma^2)$$

Para testar a hipótese nula de não ligação, poder-se-ia utilizar a teste de razão de verossimilhanças:

$$\lambda = -2 \ln \left[\frac{L(\hat{\mu}_1, \hat{\mu}_2, \hat{\sigma}^2, r = 0, 5)}{L(\hat{\mu}_1, \hat{\mu}_2, \hat{\sigma}^2, \hat{r})} \right]$$

Rejeitar-se-ia a hipótese se o valor de λ fosse grande para uma distribuição aproximada qui-quadrado com número de graus de liberdade calculados sob a hipótese nula. Porém, o modelo de mistura invalida a hipótese qui-quadrado (WU; MA; CASELLA, 2007) e o que normalmente se faz é um teste de permutação em λ , ou, sua variação, o escore LOD:

$$\text{LOD} = -\log \left[\frac{L(\hat{\mu}_1, \hat{\mu}_2, \hat{\sigma}^2, r = 0, 5)}{L(\hat{\mu}_1, \hat{\mu}_2, \hat{\sigma}^2, \hat{r})} \right]$$

O teste estatístico é avaliado sobre uma grade de valores r , $r \in [0; 0, 5]$, e o máximo escore LOD indica o valor r , da grade, mais próximo à estimativa \hat{r} de máxima verossimilhança (DOERGE; ZENG; WEIR, 1997). A desvantagem deste método é o uso de análise de marcas simples, o que necessita de testes independentes para cada marca.

2.3 Abordagem bayesiana *shrinkage*

Com o advento dos SNPs, o mapeamento de QTL tornou-se um típico problema de seleção de variáveis em um modelo supersaturado. Uma alternativa para analisar este problema em modelos é usar a estimação *shrinkage*, onde todas as variáveis candidatas são incluídas no modelo, mas seus efeitos estimados são forçados a encolher para zero. Em contraste com a estimação *shrinkage* usual, onde todos os efeitos do modelo são encolhidos pelo mesmo fator, a inferência bayesiana permite atribuir prioris para cada efeito de marcadores e fazer o fator *shrinkage* variar de acordo com esses diferentes efeitos (WANG et al., 2005). Este método *shrinkage* força intervalos de marcadores que não contêm QTLs a ter seus efeitos estimados próximos de zero, enquanto que intervalos contendo QTLs sofrem pouco encolhimento nos efeitos estimados.

Isto tem evitado os problemas de seleção de modelos, mas como todos os possíveis QTLs (mesmo os de efeitos nulos) são mantidos no modelo, a ideia de modelo parcimonioso parece ser violada, exigindo alta demanda computacional. Contudo, esta análise é capaz de penalizar os pequenos efeitos e dar-lhes pesos insignificantes para que suas inclusões tenham um efeito desprezível sobre a análise. Uma das boas propriedades da análise bayesiana é sua habilidade de tratar modelos com número grande de variáveis (XU, 2003).

Além disso, em comparação com outros métodos de análise, esta abordagem tem se mostrado superior tanto na estimativa dos efeitos quanto na estimativa da real posição do QTL (WANG et al., 2005). Apesar disso, ainda poucos estudos de mapeamento de QTL são realizados utilizando-a, principalmente, em virtude da alta demanda computacional.

Um resultado encontrado por Xu (2003), na análise bayesiana de múltiplas marcas, é que marcadores de grandes efeitos nas análises por marcas individuais

mantêm seus grandes efeitos na análise bayesiana, enquanto marcadores com pequenos efeitos têm sido "encolhidos". Este fator de encolhimento refere-se à penalização de cada possível QTL pela razão da sua variância com a variância residual. Assim, aqueles QTLs de pequeno efeito e baixa variância têm esses efeitos "encolhidos" a valores próximos a zero, pois são penalizados pela variância residual do modelo. Por outro lado, aqueles QTLs de grande efeito tendem a apresentar destacada variância e são menos penalizados pela variância residual (BALESTRE et al., 2012).

3 MATERIAL E MÉTODOS

Utilizou-se o *software* R (R CORE TEAM, 2014) para realizar as análises e obter os gráficos.

3.1 Simulação

Foram simulados 300 indivíduos de uma população F_2 , com herdabilidade 0,5, utilizando o programa QGene (JOEHANES; NELSON, 2008). Para essa população, criaram-se onze grupos de ligação com tamanho de 120cM cada e distância média de 10cM no genoma, totalizando 165 marcadores SNPs (*Single-Nucleotide Polymorphism*). Sete supostos QTLs foram distribuídos, aleatoriamente, nos grupos de ligação, com seus respectivos efeitos aditivos e dominantes.

Na Tabela 1 abaixo mostra-se o grupo de ligação em que o QTL simulado está alocado, o valor da sua posição neste grupo, em centiMorgan (cM), seus efeitos aditivo e dominante. Na Figura 1 estão representados os grupos de ligação, as marcas e os QTLs simulados.

Tabela 1 Posições, efeitos aditivos e efeitos dominantes dos QTLs simulados nos respectivos grupos de ligação.

| | Grupo de Ligação | Posição | Efeito Aditivo | Efeito Dominante |
|-------|------------------|---------|----------------|------------------|
| QTL 1 | 1 | 76,6 | -10 | 25 |
| QTL 2 | 1 | 102,1 | 5 | 20 |
| QTL 3 | 5 | 22,3 | -3 | 9 |
| QTL 4 | 7 | 18,7 | -7 | 5 |
| QTL 5 | 7 | 96,5 | 15 | -3 |
| QTL 6 | 8 | 50,6 | 10 | 8 |
| QTL 7 | 9 | 52,8 | 20 | 5 |

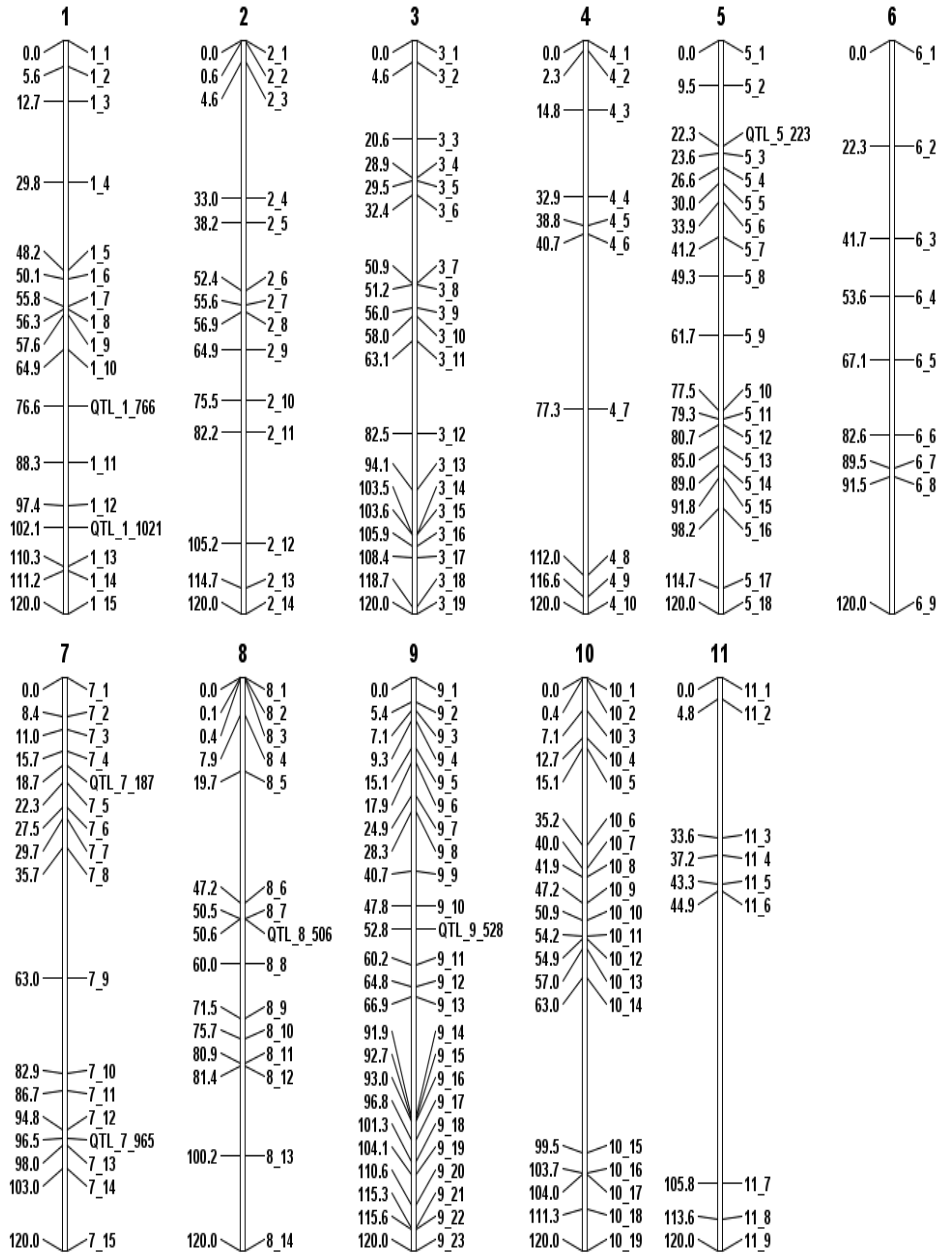


Figura 1 Mapa genético simulado com onze grupos de ligação, 165 marcadores e sete QTLs.

3.2 Mapeamento por múltiplos marcadores

Este método foi proposto por Xu (2003), sob uma abordagem bayesiana, onde cada marcador é considerado um provável QTL, fazendo com que a matriz \mathbf{X} seja totalmente observável.

3.2.1 Modelo Linear

Sejam y_i , para $i = 1, \dots, n$, os valores fenotípicos do indivíduo i em uma população F_2 . O modelo linear é:

$$y_i = b_0 + \sum_{j=1}^p x_{ij}b_j + \sum_{j=1}^p w_{ij}d_j + e_i \quad (1)$$

sendo

- b_0 a média geral da população sob estudo;
- p o número total de QTLs assumido - neste caso, é o número de marcadores;
- x_{ij} a variável representativa indicando o efeito aditivo do marcador j para o indivíduo i ;
- w_{ij} a variável representativa indicando o efeito de dominância do marcador j para o indivíduo i ;
- b_j o efeito aditivo do QTL associado com o marcador j ;
- d_j o efeito de dominância do QTL associado com o marcador j ;
- e_i o resíduo com distribuição $N(0, \sigma_0^2)$.

As variáveis x_{ij} e w_{ij} são definidas como:

$$x_{ij} = \begin{cases} 1, & \text{para } AA; \\ 0, & \text{para } Aa; \\ -1, & \text{para } aa. \end{cases} \quad w_{ij} = \begin{cases} -1/2, & \text{para } AA; \\ 1/2, & \text{para } Aa; \\ -1/2, & \text{para } aa. \end{cases} \quad (2)$$

3.2.2 Estimação Bayesiana

Neste modelo, as variáveis observadas são os dados fenotípicos $\mathbf{y} = \{y_i\}$, para $i = 1, \dots, n$, e os genótipos dos marcadores, enquanto as variáveis não observadas são a média, os efeitos aditivos e dominantes, representados por $\mathbf{c} = \{b_0, b_j, d_j\}$ e suas variâncias $\mathbf{v} = \{\sigma_0^2, \sigma_{b_j}^2, \sigma_{d_j}^2\}$, $j = 1, \dots, p$.

As distribuições a priori são:

$$p(b_0) \propto 1, \quad p(\sigma_0^2) \propto \frac{1}{\sigma_0^2}, \quad p(b_j) \propto N(0, \sigma_{b_j}^2), \\ p(d_j) \propto N(0, \sigma_{d_j}^2), \quad p(\sigma_{b_j}^2) \propto \frac{1}{\sigma_{b_j}^2}, \quad p(\sigma_{d_j}^2) \propto \frac{1}{\sigma_{d_j}^2}, \quad j = 1, \dots, p$$

Assumindo independência, a priori conjunta das variáveis não observadas $\mathbf{p}(\mathbf{c}, \mathbf{v})$ é o produto das prioris dos parâmetros individuais. Para o modelo descrito em (1), a verossimilhança pode ser dada pela distribuição Normal Multivariada descrita abaixo:

$$p(\mathbf{y}|\mathbf{c}, \mathbf{v}) = \prod_{i=1}^n p(y_i|\mathbf{c}, \sigma_0^2) \propto \\ \propto (\sigma_0^2)^{-n/2} \exp \left\{ -\frac{1}{2\sigma_0^2} \sum_{i=1}^n \left(y_i - b_0 - \sum_{j=1}^p x_{ij}b_j - \sum_{j=1}^p w_{ij}d_j \right)^2 \right\}$$

A distribuição conjunta a posteriori pode ser representada por:

$$p(\mathbf{c}, \mathbf{v} | \mathbf{y}) \propto p(\mathbf{y} | \mathbf{c}, \mathbf{v}) p(\mathbf{c}, \mathbf{v})$$

3.2.3 Amostragem das distribuições condicionais completas

Utiliza-se o algoritmo MCMC (Monte Carlo via Cadeia de Markov), baseado no amostrador de Gibbs, que usa as distribuições condicionais completas a posteriori para amostrar valores dos parâmetros e obter a distribuição conjunta a posteriori. O amostrador é realizado de acordo com a sequência abaixo.

Inicialização: Os parâmetros b_0 e σ_0^2 são inicializados com a média e a variância dos dados fenotípicos, respectivamente; b_j e d_j são inicializados com o valor zero e $\sigma_{b_j}^2$ e $\sigma_{d_j}^2$ são inicializadas com o valor 0,5.

$$\mathbf{I}^{(k)} = \left[b_0^{(k)}, \dots, b_p^{(k)}, \sigma_0^{2(k)}, \dots, \sigma_p^{2(k)} \right]$$

sendo k o número da iteração atual, iniciando em zero.

Atualizar b_0 : A distribuição condicional a posteriori para b_0 é Normal com média \bar{b}_0 e variância s_0^2 , apresentadas a seguir. O b_0 amostrado é denotado por $b_0^{(k+1)}$ e colocado no lugar de $b_0^{(k)}$ em todos os processos de amostragem subsequentes.

$$\bar{b}_0 = \frac{1}{n} \sum_{i=1}^n \left(y_i - \sum_{j=1}^p x_{ij} b_j^{(k)} - \sum_{j=1}^p w_{ij} d_j^{(k)} \right)$$

$$s_0^2 = \frac{1}{n} \sigma_0^{2(k)}$$

Atualizar b_j : A distribuição condicional a posteriori para b_j é Normal com média \bar{b}_j e variância $s_{b_j}^2$, descritas abaixo. O b_j amostrado é denotado por $b_j^{(k+1)}$ e colocado no lugar de $b_j^{(k)}$.

$$\bar{b}_j = \left(\sum_{i=1}^n x_{ij}^2 + \frac{\sigma_0^2(k)}{\sigma_{b_j}^2(k)} \right)^{-1} \sum_{i=1}^n x_{ij} \left(y_i - b_0^{(k)} - \sum_{t \neq j}^p x_{it} b_t^{(k)} - \sum_{j=1}^p w_{ij} d_j^{(k)} \right)$$

$$s_{b_j}^2 = \left(\sum_{i=1}^n x_{ij}^2 + \frac{\sigma_0^2(k)}{\sigma_{b_j}^2(k)} \right)^{-1} \sigma_0^2(k)$$

Atualizar d_j : A distribuição condicional a posteriori para d_j é Normal com média \bar{d}_j e variância $s_{d_j}^2$, descritas abaixo. Os d_j amostrados são denotados por $d_j^{(k+1)}$ e colocados no lugar de $d_j^{(k)}$.

$$\bar{d}_j = \left(\sum_{i=1}^n w_{ij}^2 + \frac{\sigma_0^2(k)}{\sigma_{d_j}^2(k)} \right)^{-1} \sum_{i=1}^n w_{ij} \left(y_i - b_0^{(k)} - \sum_{j=1}^p x_{ij} b_j^{(k)} - \sum_{t \neq j}^p w_{it} d_t^{(k)} \right)$$

$$s_{d_j}^2 = \left(\sum_{i=1}^n w_{ij}^2 + \frac{\sigma_0^2(k)}{\sigma_{d_j}^2(k)} \right)^{-1} \sigma_0^2(k)$$

Atualizar σ_0^2 : A distribuição condicional a posteriori para a variância residual é Qui-Quadrado Inversa Escalada. A variância amostrada $\sigma_0^2(k+1)$ é colocada no lugar de $\sigma_0^2(k)$.

$$p(\sigma_0^2 | \dots) \sim \chi_{esc}^{-2}(n, FQ)$$

sendo

$$FQ = \sum_{i=1}^n \left(y_i - b_0^{(k)} - \sum_{j=1}^p x_{ij} b_j^{(k)} - \sum_{j=1}^p w_{ij} d_j^{(k)} \right)^2$$

Atualizar $\sigma_{b_j}^2$ e $\sigma_{d_j}^2$: As distribuições condicionais a posteriori para $\sigma_{b_j}^2$ e $\sigma_{d_j}^2$ são distribuições Qui-Quadrado Inversas Escaladas. As variâncias amostradas $\sigma_{b_j}^{2(k+1)}$ e $\sigma_{d_j}^{2(k+1)}$ são colocadas no lugar de $\sigma_{b_j}^{2(k)}$ e $\sigma_{d_j}^{2(k)}$, respectivamente.

$$p\left(\sigma_{b_j}^2 | \dots\right) \sim \chi_{esc}^{-2}\left(1, b_j^{2(k)}\right) \quad p\left(\sigma_{d_j}^2 | \dots\right) \sim \chi_{esc}^{-2}\left(1, d_j^{2(k)}\right)$$

Repete-se a sequência acima até a convergência da cadeia para uma distribuição estacionária. Obtida a convergência, os valores finais amostrados são utilizados para a estimação dos parâmetros.

3.3 *Moving away from the marker* com análise bayesiana

A técnica de Doerge, Zeng e Weir (1997) apresentada na seção 2.2 pode ser avaliada sob abordagem bayesiana, incluindo uma adaptação da análise em múltiplos marcadores de Xu (2003) e da análise de múltiplos QTLs de Wang et al. (2005). Esta técnica adaptada permite que o QTL assuma posições variando dentro de um intervalo (distância) definido pela fração de recombinação entre o QTL e o marcador, em vez de fixo entre duas marcas, como no mapeamento por múltiplos QTLs, adotando-se uma busca mais restrita.

Neste trabalho, utilizou-se o intervalo definido pelas frações de recombinação zero e 0,2, adotando uma busca de QTLs à esquerda e à direita das marcas dentro deste intervalo. Assim, assume-se que a marca está ligada ao QTL e, então, caminha-se dentro do intervalo escolhido, afastando-se da marca que está sendo utilizada como pivô. Portanto, o provável QTL poderá localizar-se em alguma posição dentro deste intervalo.

3.3.1 Modelo Linear

O modelo linear adotado, para $y_i, i = 1, \dots, n$, sendo o valor fenotípico do indivíduo i em uma população F_2 , é:

$$y_i = b_0 + \sum_{j=1}^p x_{ij}b_j + \sum_{j=1}^p w_{ij}d_j + e_i \quad (3)$$

sendo

- b_0 a média geral da população sob estudo;
- p o número total de QTLs assumido - neste caso, é o número de marcadores;
- x_{ij} a variável representativa indicando o efeito aditivo do QTL j para o indivíduo i ;
- w_{ij} a variável representativa indicando o efeito de dominância do QTL j para o indivíduo i ;
- b_j o efeito aditivo do QTL j ;
- d_j o efeito de dominância do QTL j ;
- e_i o resíduo com distribuição $N(0, \sigma_0^2)$.

As variáveis x_{ij} e w_{ij} são definidas como em (2). Considere $\mathbf{m} = m_{ij}, i = 1, \dots, n$ e $j = 1, \dots, p$, o vetor de marcadores, sendo o marcador j relacionado ao indivíduo i .

3.3.2 Estimação Bayesiana

As variáveis observadas são os dados fenotípicos $\mathbf{y} = \{y_i\}, i = 1, \dots, n$ observações, e os genótipos dos marcadores, enquanto as variáveis não observadas

são as posições dos QTLs $\lambda = \{\lambda_j\}$, a média, os efeitos dos QTLs, representados por $\mathbf{c} = \{b_0, b_j, d_j\}$, suas variâncias $\mathbf{v} = \{\sigma_0^2, \sigma_{b_j}^2, \sigma_{d_j}^2\}$ e as variáveis representativas dos genótipos dos QTLs x_{ij} e $w_{ij}, j = 1, \dots, p$.

As prioris assumidas são:

$$p(b_0) \propto 1, \quad p(\sigma_0^2) \propto \frac{1}{\sigma_0^2}, \quad p(b_j) = N(0, \sigma_{b_j}^2)$$

$$p(d_j) = N(0, \sigma_{d_j}^2), \quad p(\sigma_{b_j}^2) \propto \frac{1}{\sigma_{b_j}^2}, \quad p(\sigma_{d_j}^2) \propto \frac{1}{\sigma_{d_j}^2}, \quad j = 1, \dots, p$$

Os genótipos x_{ij} e w_{ij} são não observados, mas podem ser inferidos com base na informação do marcador j e das posições λ_j s dos QTLs em relação ao marcador j .

$$p(\mathbf{x}, \mathbf{w} | \boldsymbol{\lambda}, \mathbf{m}) = p(\boldsymbol{\lambda}, \mathbf{m} | \mathbf{x}, \mathbf{w}) p(\mathbf{x}, \mathbf{w})$$

sendo

$$p(\mathbf{x}, \mathbf{w}) = \begin{cases} 1/4, & \text{para } AA; \\ 1/2, & \text{para } Aa; \\ 1/4, & \text{para } aa. \end{cases}$$

$$p(\boldsymbol{\lambda}, \mathbf{m} | \mathbf{x}, \mathbf{w}) = \begin{bmatrix} \frac{1}{4}(1-r)^2 & \frac{1}{2}r(1-r) & \frac{1}{4}r^2 \\ \frac{1}{2}r(1-r) & \frac{1}{2}[(1-r)^2 + r^2] & \frac{1}{2}r(1-r) \\ \frac{1}{4}r^2 & \frac{1}{2}(1-r) & \frac{1}{4}(1-r)^2 \end{bmatrix}$$

Nesse estudo, fixando-se a marca, o QTL pode variar sua posição dentro do intervalo especificado acima, sendo mais ligado ou menos ligado à marca; deste modo, a priori, λ_j é uniformemente distribuída em um intervalo correspondente à fração de recombinação variando de zero (marca é o próprio QTL) a 0,2 (a marca

e o QTL estão pouco ligados):

$$p(\lambda_j) = \frac{1}{0,2}$$

A priori conjunta das variáveis não observáveis $p(\mathbf{c}, \mathbf{v})$, assumindo independência, é o produto das prioris dos parâmetros individuais.

$$p(\mathbf{c}, \mathbf{v}, \mathbf{x}, \mathbf{w}, \boldsymbol{\lambda}) = p(b_0)p(\sigma_0^2)p(\mathbf{x}, \mathbf{w}|\boldsymbol{\lambda}, \mathbf{m}) \prod_{j=1}^p p(b_j)p(\sigma_{b_j}^2)p(d_j)p(\sigma_{d_j}^2)p(\lambda_j)$$

A verossimilhança é descrita como:

$$p(\mathbf{y}, \mathbf{m}|\mathbf{c}, \mathbf{v}, \mathbf{x}, \mathbf{w}, \boldsymbol{\lambda}) = \prod_{i=1}^n p(y_i, \mathbf{m}|\mathbf{c}, \mathbf{x}, \mathbf{w}, \sigma_0^2)$$

$$\propto (\sigma_0^2)^{-n/2} \exp \left\{ -\frac{1}{2\sigma_0^2} \sum_{i=1}^n \left(y_i - b_0 - \sum_{j=1}^p x_{ij}b_j - \sum_{j=1}^p w_{ij}d_j \right)^2 \right\}$$

A distribuição a posteriori é:

$$p(\mathbf{c}, \mathbf{v}, \mathbf{x}, \mathbf{w}, \boldsymbol{\lambda}|\mathbf{y}, \mathbf{m}) \propto p(\mathbf{y}, \mathbf{m}|\mathbf{c}, \mathbf{v}, \mathbf{x}, \mathbf{w}, \boldsymbol{\lambda})p(\mathbf{c}, \mathbf{v}, \mathbf{x}, \mathbf{w}, \boldsymbol{\lambda})$$

3.3.3 Amostragem das distribuições condicionais completas

Da mesma forma que no item 3.2, utiliza-se o algoritmo MCMC (Monte Carlo via Cadeia de Markov), baseado na amostragem de Gibbs.

Inicialização: Os parâmetros b_0 e σ_0^2 são inicializados com a média e a variância dos dados fenotípicos, respectivamente; b_j e d_j são inicializados com o valor zero e $\sigma_{b_j}^2$ e $\sigma_{d_j}^2$ são inicializadas com valor 0,5; o parâmetro λ_j é inicia-

lizado com um valor aleatório entre zero e 0, 2; os valores iniciais de x_{ij} e w_{ij} são amostrados das probabilidades de x_{ij} e w_{ij} condicionais ao parâmetro λ_j e à marca j .

$$\mathbf{I}^{(k)} = \left[b_0^{(k)}, \dots, b_p^{(k)}, \sigma_0^{2(k)}, \dots, \sigma_p^{2(k)}, x_{ij}^{(k)}, w_{ij}^{(k)} \right]$$

sendo k o número da iteração atual, iniciando em zero.

Atualizar b_0 : A distribuição condicional a posteriori para b_0 é Normal com média \bar{b}_0 e variância σ_0^2 , apresentadas a seguir. O b_0 amostrado é denotado por $b_0^{(k+1)}$ e colocado no lugar de $b_0^{(k)}$ em todos os processos de amostragem subsequentes.

$$\bar{b}_0 = \frac{1}{n} \sum_{i=1}^n \left(y_i - \sum_{j=1}^p x_{ij}^{(k)} b_j^{(k)} - \sum_{j=1}^p w_{ij}^{(k)} d_j^{(k)} \right)$$

$$s_0^2 = \frac{1}{n} \sigma_0^{2(k)}$$

Atualizar b_j : A distribuição condicional a posteriori para b_j é Normal com média \bar{b}_j e variância $\sigma_{b_j}^2$, apresentadas abaixo. Os b_j amostrados são denotados por $b_j^{(k+1)}$ e colocados no lugar de $b_j^{(k)}$.

$$\bar{b}_j = \left(\sum_{i=1}^n x_{ij}^{2(k)} + \frac{\sigma_0^{2(k)}}{\sigma_{b_j}^2} \right)^{-1} \sum_{i=1}^n x_{ij}^{(k)} \left(y_i - b_0^{(k)} - \sum_{t \neq j}^p x_{it}^{(k)} b_t^{(k)} - \sum_{j=1}^p w_{ij}^{(k)} d_j^{(k)} \right)$$

$$s_{b_j}^2 = \left(\sum_{i=1}^n x_{ij}^{2(k)} + \frac{\sigma_0^{2(k)}}{\sigma_{b_j}^2} \right)^{-1} \sigma_0^{2(k)}$$

Atualizar d_j : A distribuição condicional a posteriori para d_j é Normal com média \bar{d}_j e variância $\sigma_{d_j}^2$, apresentadas abaixo. Os d_j amostrados são denotados por $d_j^{(k+1)}$ e colocados no lugar de $d_j^{(k)}$.

$$\bar{d}_j = \left(\sum_{i=1}^n w_{ij}^2{}^{(k)} + \frac{\sigma_0^2{}^{(k)}}{\sigma_{d_j}^2{}^{(k)}} \right)^{-1} \sum_{i=1}^n w_{ij}^{(k)} \left(y_i - b_0^{(k)} - \sum_{j=1}^p x_{ij}^{(k)} b_j^{(k)} - \sum_{t \neq j}^p w_{it}^{(k)} d_t^{(k)} \right)$$

$$s_{d_j}^2 = \left(\sum_{i=1}^n w_{ij}^2{}^{(k)} + \frac{\sigma_0^2{}^{(k)}}{\sigma_{d_j}^2{}^{(k)}} \right)^{-1} \sigma_0^2{}^{(k)}$$

Atualizar σ_0^2 : A distribuição condicional a posteriori para a variância residual é Qui-Quadrado Inversa Escalada. A variância amostrada $\sigma_0^2{}^{(k+1)}$ é colocada no lugar de $\sigma_0^2{}^{(k)}$.

$$p(\sigma_0^2 | \dots) \sim \chi_{esc}^{-2}(n, FQ)$$

sendo

$$FQ = \sum_{i=1}^n \left(y_i - b_0^{(k)} - \sum_{j=1}^p x_{ij}^{(k)} b_j^{(k)} - \sum_{j=1}^p w_{ij}^{(k)} d_j^{(k)} \right)^2$$

Atualizar $\sigma_{b_j}^2$ e $\sigma_{d_j}^2$: As distribuições condicionais a posteriori para $\sigma_{b_j}^2$ e $\sigma_{d_j}^2$ são distribuições Qui-Quadrado Inversas Escaladas. As variâncias amostradas $\sigma_{b_j}^2{}^{(k+1)}$ e $\sigma_{d_j}^2{}^{(k+1)}$ são colocadas no lugar de $\sigma_{b_j}^2{}^{(k)}$ e $\sigma_{d_j}^2{}^{(k)}$, respectivamente.

$$p(\sigma_{b_j}^2 | \dots) \sim \chi_{esc}^{-2}(1, b_j^2{}^{(k)}) \quad p(\sigma_{d_j}^2 | \dots) \sim \chi_{esc}^{-2}(1, d_j^2{}^{(k)})$$

Atualizar x_{ij} e w_{ij} : Cada genótipo do QTL amostrado tem distribuição Bernoulli, usando a informação da marca j , com probabilidade dada abaixo.

$$\begin{aligned}
& p\left(x_{ij}, w_{ij} | \lambda_j^{(k)}, m_{ij}, y_i, \mathbf{c}^{(k)}, \sigma_0^{2(k)}\right) = \\
& = \frac{p\left(x_{ij}, w_{ij} | \lambda_j^{(k)}, m_{ij}\right) p\left(y_i | \mathbf{c}^{(k)}, x_{ij}, w_{ij}, \sigma_0^{2(k)}\right)}{\sum_{l,h} p(x_{ij} = l, w_{ij} = h | \lambda_j, m_j) p(y_i | \mathbf{c}^{(k)}, x_{ij} = l, w_{ij} = h, \sigma_0^2)}
\end{aligned}$$

sendo l e h os genótipos apresentados em (2).

Atualizar λ_j : O parâmetro λ_j é difícil de amostrar diretamente de sua distribuição condicional a posteriori sendo, portanto, necessário usar um passo do algoritmo Metropolis-Hastings (HASTINGS, 1970; METROPOLIS et al., 1953) dentro do algoritmo de amostragem de Gibbs. No método apresentado, pode-se utilizar uma distribuição uniforme como função auxiliar, onde se amostra uma nova posição ($\lambda_j^{(k+1)}$), utilizando a função de Haldane (HALDANE, 1919), sob um intervalo delimitado por $\max(0, \lambda_j - c)$ e $\min(0, 2; \lambda_j + c)$, onde c é uma constante que define o caminhamento dentro do intervalo j , normalmente valor de 1 ou 2 cM. Esta função é denotada por $u(\lambda_j^{(k+1)}, \lambda_j^{(k)})$, e a nova posição será aceita na k -ésima iteração com $\min(1, \alpha)$ de probabilidade. Assim, se α for aceito, uma nova posição é estabelecida e um novo genótipo é sugerido para x_{ij} e w_{ij} .

$$\alpha = \frac{p\left(\lambda_j^{(k+1)} | \mathbf{y}, \mathbf{c}^{(k)}, \sigma_0^{2(k)}, m_j, \mathbf{x}^{(k)}, \mathbf{w}^{(k)}\right) u\left(\lambda_j^{(k+1)}, \lambda_j^{(k)}\right)}{p\left(\lambda_j^{(k)} | \mathbf{y}, \mathbf{c}^{(k)}, \sigma_0^{2(k)}, m_j, \mathbf{x}^{(k)}, \mathbf{w}^{(k)}\right) u\left(\lambda_j^{(k)}, \lambda_j^{(k+1)}\right)}$$

Repete-se a sequência acima até a suposta convergência da cadeia para uma distribuição estacionária. Na cadeia final, a média a posteriori dos genótipos finais amostrados foi utilizada para a estimação dos parâmetros.

3.4 Análise pós-MCMC

Na análise bayesiana simples, a posição do QTL é inferida com base na quantidade de vezes que o efeito do QTL passa por pequena região (bin) em determinada posição do genoma. Essa curva descreve o perfil de intensidade do QTL. Na abordagem de Wang et al. (2005), é assumido que cada intervalo está associado a um QTL de modo que o suposto QTL irá passar por todas as regiões dentro dele e o mesmo número de vezes que o QTL passa ocorre independente do seu efeito. No entanto, é esperado ocorrer um verdadeiro QTL em um dado intervalo e sua posição demonstrará um pico, ao passo que, se o efeito é nulo, a distribuição dentro do intervalo é uniforme (YANG; XU, 2007).

O perfil da intensidade do QTL é representado por Yang e Xu (2007) como uma função da posição $f(\lambda)$. No entanto, $f(\lambda)$ pode não ser suficientemente informativa para inferência acerca do QTL na análise bayesiana *shrinkage*. Baseado nisso, Yang e Xu (2007) propuseram a descrição dos efeitos dos QTLs de acordo com suas formas quadráticas e ponderadas pela intensidade da posição:

$$g(\lambda) = W(\lambda)f(\lambda)$$

$$W(\lambda) = \mathbf{b}'\mathbf{V}_b^{-1}\mathbf{b} + \mathbf{d}'\mathbf{V}_d^{-1}\mathbf{d}$$

sendo \mathbf{b} e \mathbf{d} os vetores de efeitos aditivos e dominantes, respectivamente, dos QTLs, \mathbf{V}_b^{-1} e \mathbf{V}_d^{-1} as inversas das variâncias dos efeitos dos QTLs dadas por $\left(\sum_{i=1}^n x_{ij}^2 + \sigma_0^2/\sigma_{b_j}^2\right)^{-1} \sigma_0^2$, que corresponde à inversa da matriz de informação do efeito aditivo e $\left(\sum_{i=1}^n w_{ij}^2 + \sigma_0^2/\sigma_{d_j}^2\right)^{-1} \sigma_0^2$, que corresponde à inversa da matriz de informação do efeito dominante.

Esse teste equivale a um Teste de Wald na distribuição marginal dos parâmetros e segue aproximadamente uma distribuição qui-quadrado com dois graus de liberdade (YANG; XU, 2007). Este teste foi utilizado, nos dois métodos, para identificar as marcas significativas, sendo selecionadas quando o valor de $W(\lambda)$ foi maior que $\chi^2_{(0,95;2)} = 5,99$.

3.5 Dados do experimento real

Os dados utilizados para aplicação prática dos modelos foram retirados de Lara et al. (2014). São 186 progênies $F_{2:4}$ de feijão, obtidas considerando o cruzamento entre as linhagens CNFC 9506 e RP-2 e genotipadas com 59 marcadores, sendo 17 SSRs (*Simple Sequence Repeat*), 31 AFLPs (*Amplified Fragment Length Polymorphism*) e 11 SRAPs (*Sequence-Related Amplified Polymorphism*).

O experimento foi conduzido em delineamento em látice quadrado triplo (14x14), sendo a parcela representada por uma linha de um metro e inoculadas dez plantas por parcela, em campo experimental dentro da Universidade Federal de Lavras. O caráter estudado foi resistência à doença mofo-branco (*S. sclerotiorum*), atribuindo notas às progênies, de acordo com os níveis de reação à doença (de 1 - planta sem sintoma a 9 - morte da planta), por meio do teste de reação ao método *straw test*, ou teste do canudo. A análise fenotípica foi feita no nível de médias das parcelas, assumindo normalidade para o valor fenotípico.

3.6 Análise dos resultados

Ao longo dos grupos de ligação, simularam-se perdas de 20% e 80% das marcas, resultando em 35 e 136 marcas retiradas, respectivamente, em todo o genoma. Para cada nível de perda, o processo foi repetido 100 vezes. Considerando

os dados simulados, após a finalização de cada desbalanceamento e seleção das marcas significativas, como se sabe a posição, o efeito aditivo e o efeito dominante de cada QTL simulado, três resultados foram obtidos para cada método:

- 1 Para o mapeamento de múltiplas marcas: a distância entre a marca selecionada e o QTL simulado, a diferença entre o efeito aditivo da marca selecionada e o efeito aditivo do QTL simulado, e a diferença entre o efeito dominante da marca selecionada e o efeito dominante do QTL simulado.
- 2 Para a técnica *moving away from the marker* bayesiana: a distância entre o QTL estimado e o QTL simulado, a diferença entre o efeito aditivo do QTL estimado e o efeito aditivo do QTL simulado, e a diferença entre o efeito dominante do QTL estimado e o efeito dominante do QTL simulado. Dado, a priori, que esta técnica não se preocupa com a direção da busca do QTL, não se pode utilizar a sua distância em relação à marca para medir a distância relativa ao QTL simulado. Sendo assim, optou-se em calcular a distância direta entre a média a posteriori do genótipo do QTL estimado e o QTL simulado, utilizando a função de Kosambi (1944), de acordo com a fração de recombinação entre eles.

$$\text{distância} = 0,25 \log \left(\frac{1 + 2r}{1 - 2r} \right)$$

Para os dados reais, foi feito apenas o Teste de Wald para identificar quantas marcas foram significativas, em cada método.

4 RESULTADOS

Apenas por facilidade, adotaram-se, para esta seção, os termos Método I para designar o mapeamento por múltiplas marcas de Xu (2003) e Método II para a técnica *moving away from the marker* bayesiana proposta. São apresentados os resultados, para os dados simulados, apenas para os grupos de ligação que continham QTLs simulados (grupos de ligação um, cinco, sete, oito e nove).

4.1 Dados simulados

4.1.1 Perda de 20% das marcas

Retirando-se 20% das marcas em cada grupo de ligação, restaram 130 marcas no genoma. Avaliaram-se a distância, a diferença absoluta entre os efeitos aditivos e a diferença absoluta entre os efeitos dominantes do QTL simulado com o QTL estimado (ou com a marca selecionada), para os dois métodos.

Na Figura 2 mostra-se o poder dos marcadores em detectar QTL, no primeiro grupo de ligação (GL1), para o Método I (em vermelho) e para o Método II (em azul). O gráfico de poder é dado por $\tau(\frac{F}{d_m})$, sendo F o número de vezes que o QTL foi significativo dentre as seleções e d_m a sua distância média do QTL simulado. Quanto maior o número de vezes que o QTL for encontrado por uma marca e com menor distância do QTL simulado, maior será a estatística τ e, consequentemente, maior o pico do gráfico. Assim, se o QTL encontrado estiver extremamente próximo ao verdadeiro, a distância tenderá a zero e o pico tenderá ao infinito. As setas representam os QTLs simulados, entre as marcas, mas não sua magnitude (efeitos).

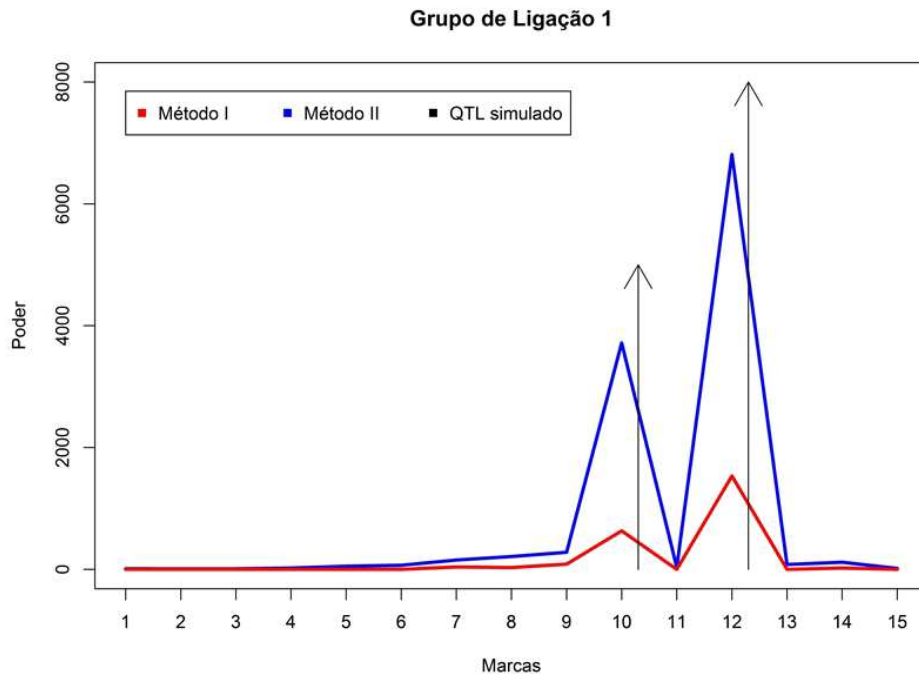


Figura 2 Poder dos Métodos I e II para detectar QTL, para o primeiro grupo de ligação simulado, com 20% de perda de marcadores.

Observa-se que ambos os métodos encontraram os dois QTLs simulados com distância muito pequena dos verdadeiros, sendo o Método II com maior poder. Por exemplo, o marcador dez é considerado um provável QTL (74 vezes significativo em 82 seleções), pelo Método I, com distância média de 11,7cM do verdadeiro QTL; já no Método II, esse marcador encontrou um QTL com distância média do verdadeiro de 1,91cM, sendo significativo 71 vezes em 82 seleções (Tabela 2, Apêndice C; Figura 4).

Para os efeitos aditivos o procedimento foi semelhante: dividiu-se o número de vezes que o QTL foi significativo, dentre as seleções, pela média das diferenças absolutas entre os efeitos aditivos do QTL encontrado e do QTL simulado. O mesmo foi feito para os efeitos dominantes. Assim, um resultado satisfatório

seria obter frequências altas de detecção e médias baixas das diferenças absolutas dos efeitos, o que resultaria em um pico alto. Na Figura 3 representa-se este procedimento para o efeito aditivo no primeiro grupo de ligação, sendo o Método I na cor vermelha e o Método II na cor azul.

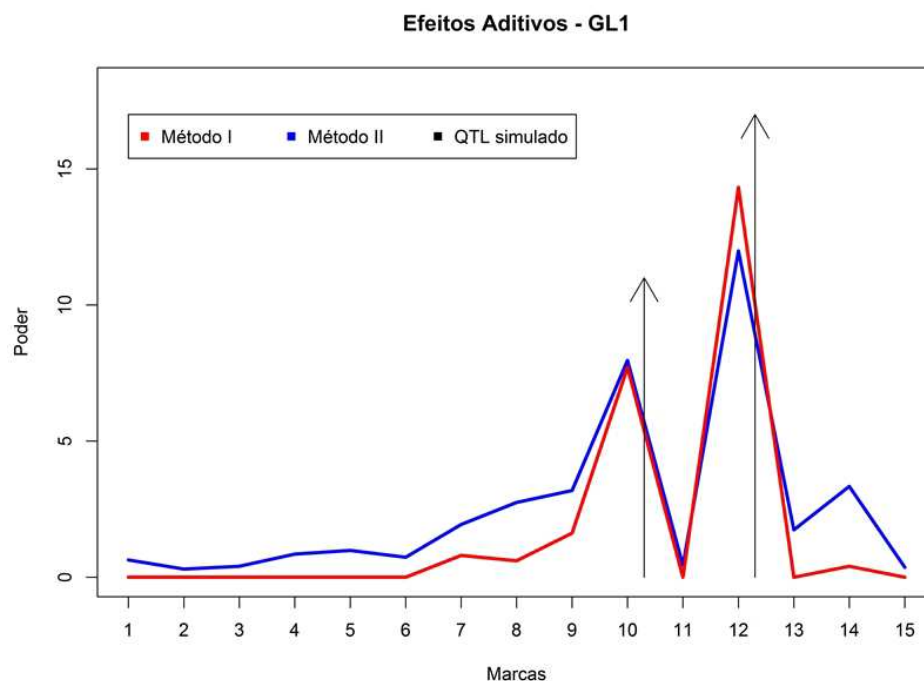


Figura 3 Poder dos Métodos I e II para determinar efeito aditivo dos QTLs, para o primeiro grupo de ligação, com 20% de perda de marcadores.

Pode-se observar que o Método II é mais poderoso para determinar efeitos aditivos com valores mais próximos do verdadeiro, sendo pouco menos poderoso que o Método I para a marca 12, semelhante ao Método I para a marca dez e mais poderoso que o Método I para as outras marcas. Para os efeitos dominantes, o Método II teve mais poder para determiná-los que o Método I, mesmo que ambos tenham sido baixos (Figura 34, Apêndice A).

Na Figura 4 resumem-se as informações acerca dos QTLs encontrados pe-

los Métodos I e II, respectivamente, para o primeiro grupo de ligação: as frequências relativas de detecção do QTL com seus respectivos intervalos de confiança exatos para proporção com 5% de significância, as posições dos QTLs encontrados pelas suas respectivas marcas, e a média da diferença absoluta dos efeitos aditivos entre o QTL encontrado e o QTL simulado. Com isso, frequências relativas altas com intervalos de confiança pequenos e diferença de efeitos pequena representam boa detecção. As setas representam os QTLs simulados, mas não os seus efeitos (magnitude); M1 representa o marcador um no Método I; M2, o marcador dois no Método I, assim por diante; Q1 representa o QTL encontrado pelo marcador um no Método II; Q2, o QTL encontrado pelo marcador dois no Método II; assim por diante; os pontos em preto no eixo x são as posições dos marcadores no grupo de ligação. A mesma ideia é usada na Figura 44, Apêndice B, mas com a representação das diferenças médias dos efeitos dominantes.

Observa-se que as marcas dez e 12 encontraram QTLs próximos aos verdadeiros em ambos os métodos, com frequências altas (Tabela 2, Apêndice C), sendo consideradas, portanto, marcas notáveis. No Método I, os QTLs estimados por essas marcas são fixos nas posições correspondentes a elas; já no Método II, os QTLs estimados por essas marcas se encontram mais próximos aos QTLs simulados (setas). A diferença de efeitos aditivos para essas marcas foi baixa (Figura 4) e de efeitos dominantes foi alta (Figura 42, Apêndice B), para os dois métodos.

Note que, no Método I, marcas posicionadas longe da região do QTL simulado têm frequências de detecção com valor zero, não identificando nenhum QTL; já no Método II, todas identificaram QTLs, mas as mais afastadas têm frequências muito baixas, evidenciando que estão realmente distantes e pouco ligadas ou não ligadas aos QTLs.

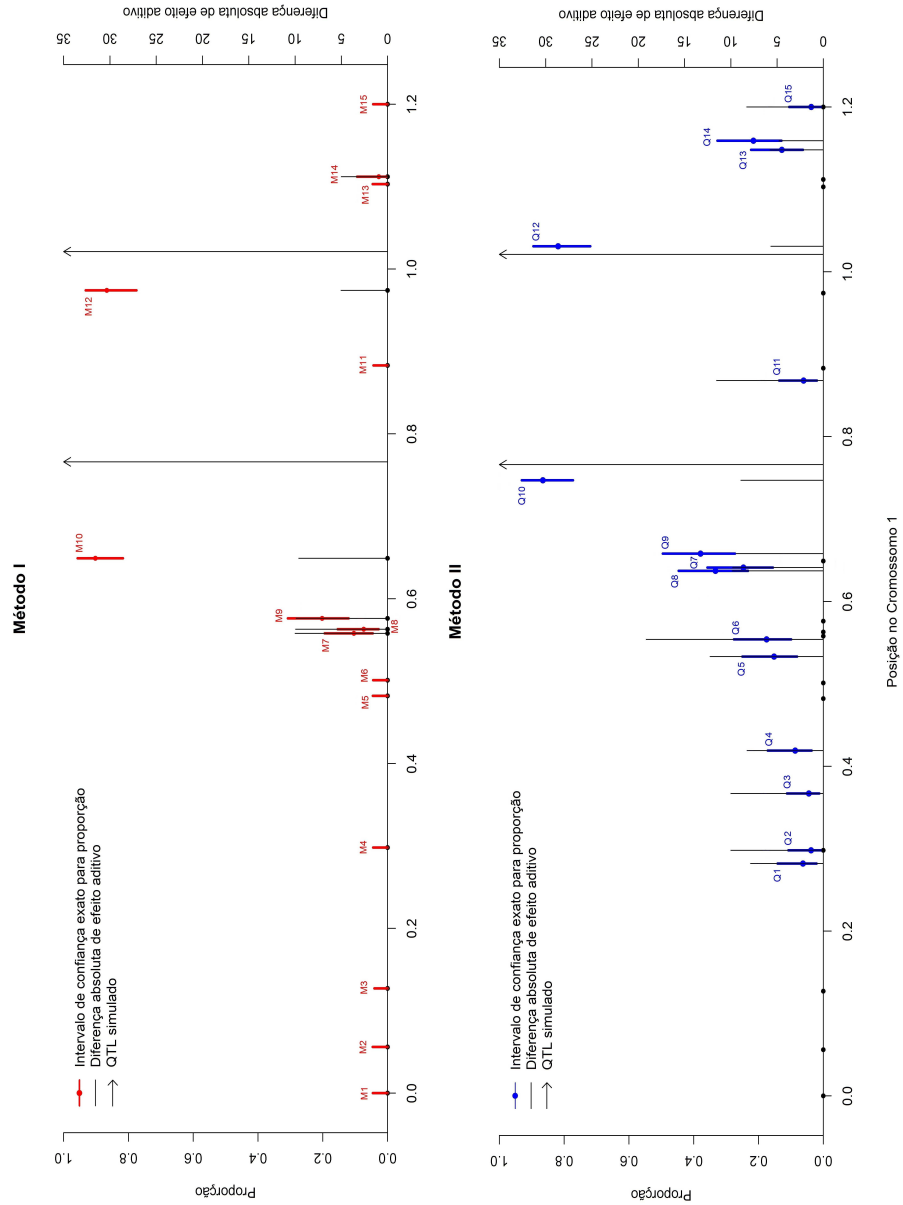


Figura 4 QTLs estimados pelos Métodos I e II com as respectivas posições (em unidade Morgan) no cromossomo um, frequências relativas com intervalos de confiança exato para proporção com 5% de significância e diferenças absolutas dos efeitos aditivos entre os QTLs estimados e os QTLs simulados, para perda de 20% das marcas.

Nas Figuras 5 e 6 mostram-se o poder de detecção de QTL e de efeitos aditivos, respectivamente, para o grupo de ligação cinco pelos dois métodos, seguindo o procedimento descrito para as Figuras 2 e 3. O poder para determinar efeitos dominantes está representado na Figura 35, Apêndice A.

Observa-se, pela Figura 5, que o Método II tem maior poder de detecção do QTL que o Método I e com maior poder para determinar os efeitos aditivos (Figura 6). O marcador dois é um marcador notável, já que foi significativo 40 vezes em 74 seleções pelo Método II e 37 vezes em 74 seleções pelo Método I. Para determinar os efeitos dominantes, o Método II, também, foi mais poderoso que o Método I (Figura 35, Apêndice A).

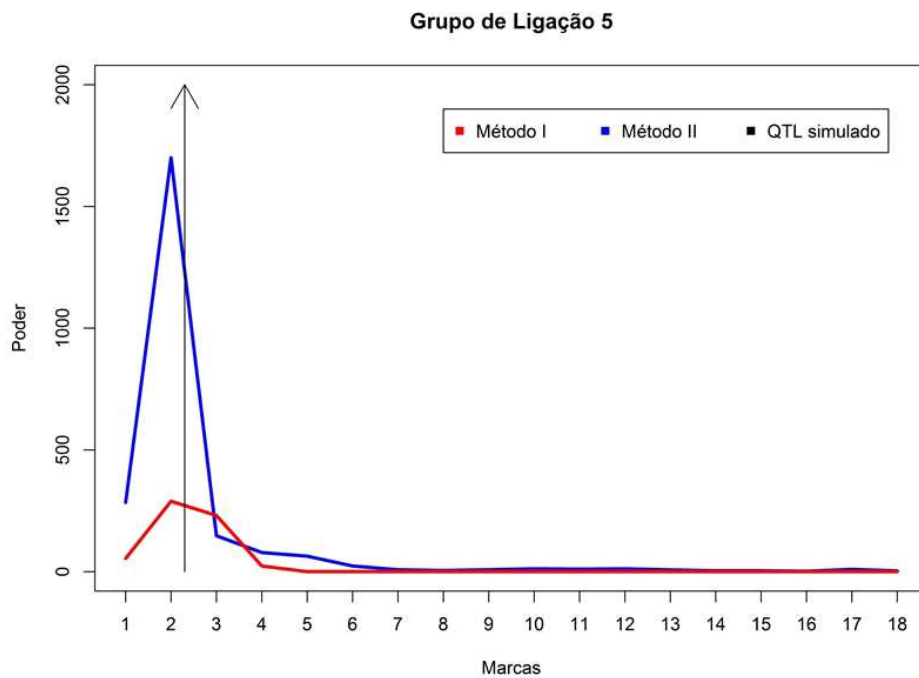


Figura 5 Poder dos Métodos I e II para detectar QTL, para o quinto grupo de ligação simulado, com 20% de perda de marcadores.

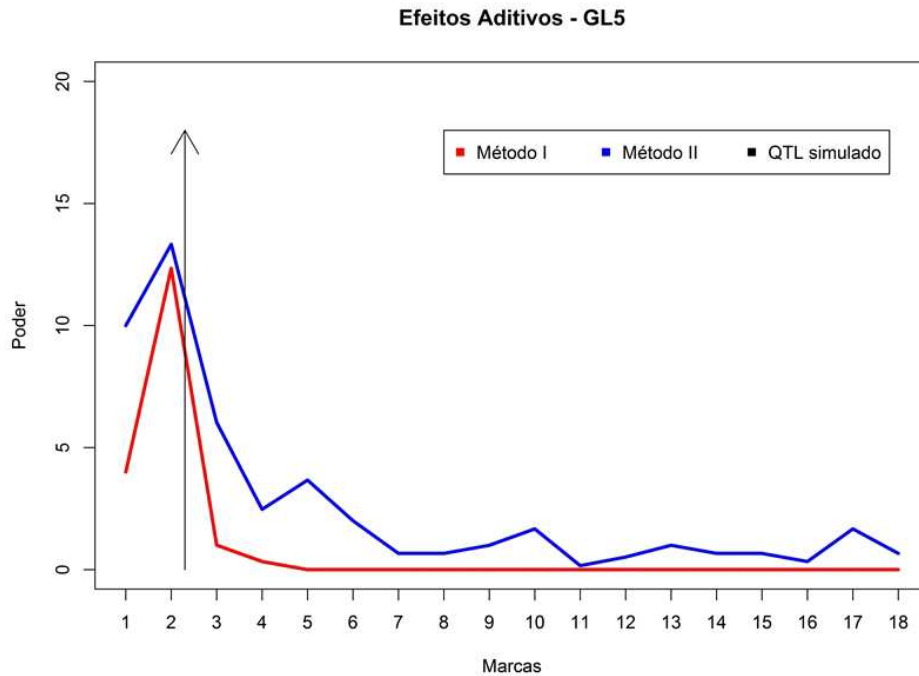


Figura 6 Poder dos Métodos I e II para determinar efeito aditivo dos QTLs, para o quinto grupo de ligação, com 20% de perda de marcadores.

Na Figura 7 resumem-se as principais características acerca do QTL, para ambos os métodos, assim como as apresentadas para o primeiro grupo de ligação. Note que o marcador dois identificou o QTL simulado nos dois métodos, mas o QTL estimado, por ele, pelo Método II, está mais próximo do verdadeiro, resultando em poder de detecção maior, como visto na Figura 5. As diferenças de efeitos aditivos foram baixas em ambos os métodos (Figura 7), assim como as diferenças de efeitos dominantes (Figura 45, Apêndice B).

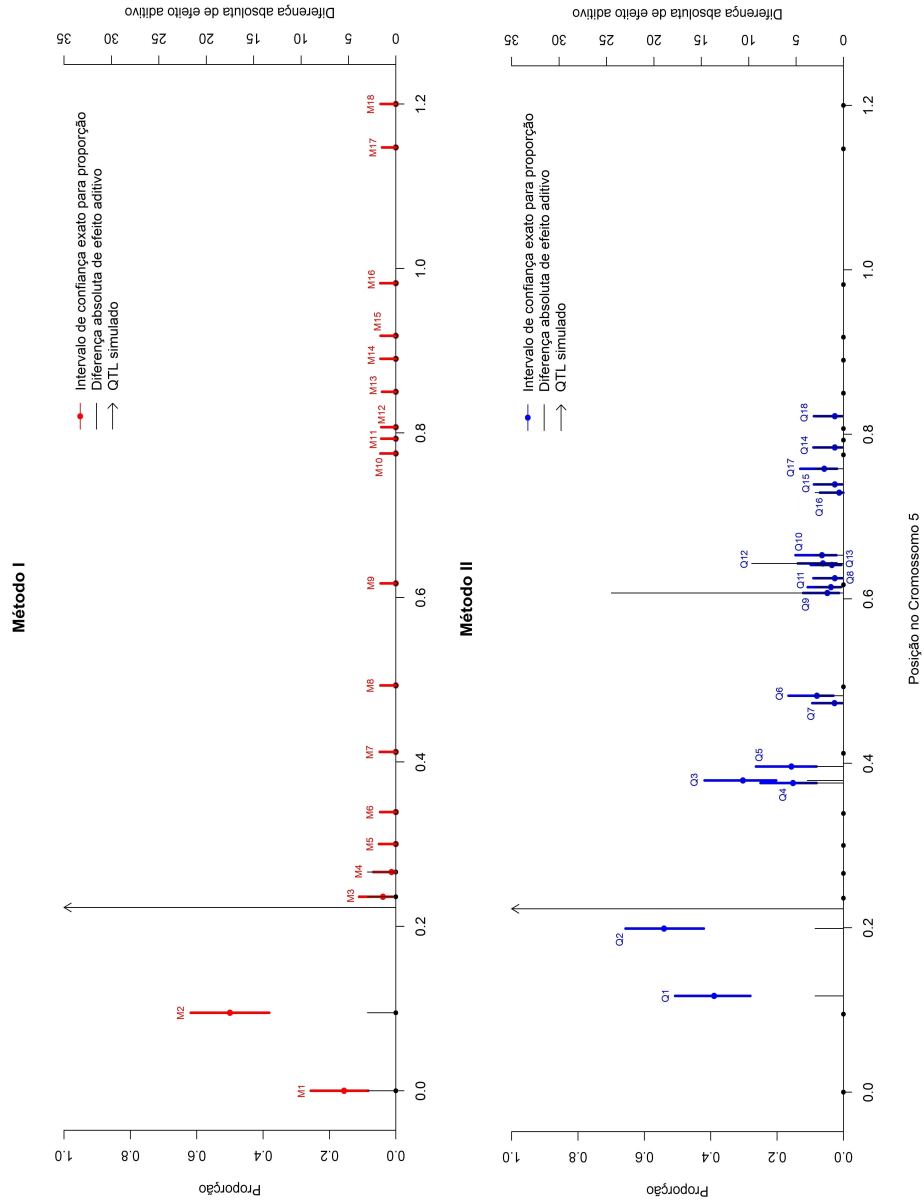


Figura 7 QTLs estimados pelos Métodos I e II com as respectivas posições (em unidade Morgan) no cromossomo cinco, frequências relativas com intervalos de confiança exato para proporção com 5% de significância e diferenças absolutas dos efeitos aditivos entre os QTLs estimados e os QTLs simulados, para perda de 20% das marcas.

Na Figura 8 representa-se o poder de detecção de ambos os métodos para o grupo de ligação sete. De acordo com essa figura, o Método I apresentou resultado melhor que o Método II para o marcador doze. Isso se deve ao fato de que a distância entre a marca e o QTL simulado é muito pequena (1,7cM), o que influenciou a altura do pico. Como o QTL foi simulado muito próximo à marca, ele foi bastante identificado pelo Método I. No Método II, como o provável QTL é analisado em cada posição dentro do intervalo estabelecido, ele pode ser estimado em qualquer uma delas. As distâncias entre o QTL estimado e o QTL simulado variaram de $6,8e^{-11}$ cM a 34cM nas 100 repetições realizadas na simulação, o que influenciou a distância média (final), resultando em 8,2cM. Com isso, o pico para a marca 12 foi baixo, mesmo que a frequência de detecção tenha sido alta e próxima à frequência obtida pelo Método I (Tabela 2, Apêndice C). Vale ressaltar que em 28 das 59 vezes em que a marca 12 foi significativa para detecção de QTL, ela estimou um QTL com distância de $6,8e^{-11}$ cM do verdadeiro, mostrando que realmente encontrou a posição do QTL simulado.

Com a marca quatro, a altura do pico na Figura 8, para o Método II, deveria ter sido maior, já que a frequência de detecção foi bem maior que pelo Método I (39 em 85 seleções pelo Método II e sete em 85 seleções pelo Método I), porém ela foi influenciada pela distância média de 9cM, maior que a distância de 3cM do Método I. Como a distância pelo Método I foi pequena, a altura do pico foi maior do que deveria, já que a frequência de detecção foi baixa. Assim, o gráfico da Figura 8 não está representando o que aconteceu, distorcendo as interpretações. Vale ressaltar que em 17 das 39 vezes em que foi significativa, a marca estimou QTL com distância de $4,13e^{-11}$ cM do verdadeiro, indicando que realmente identificou a posição do QTL simulado. Porém, as distâncias encontradas nas 85 seleções variaram de $4,13e^{-11}$ cM a 29cM, o que influenciou a distância média.

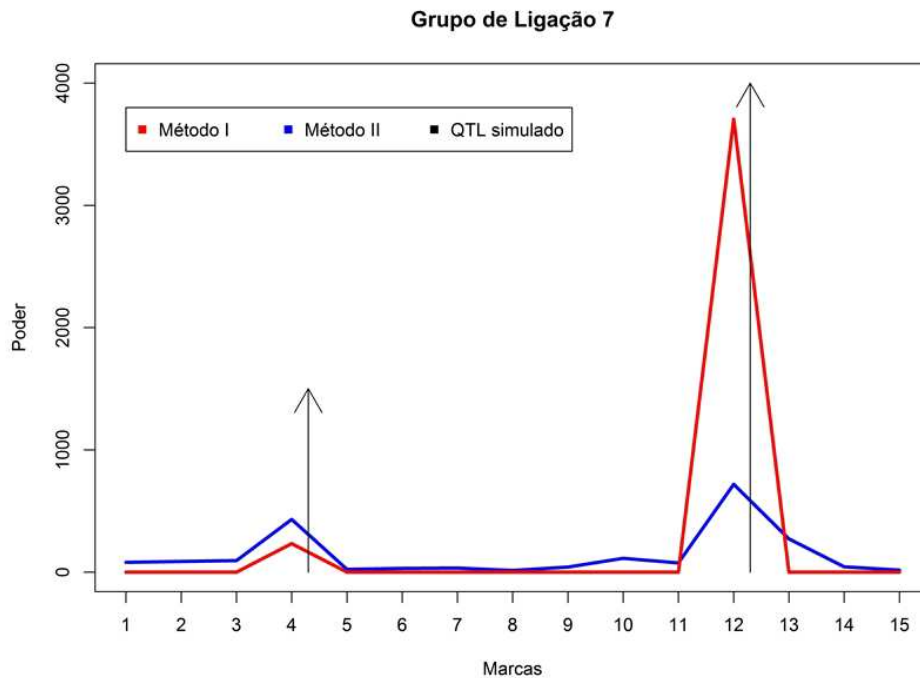


Figura 8 Poder dos Métodos I e II para detectar QTL, para o sétimo grupo de ligação simulado, com 20% de perda de marcadores.

A Figura 9 mostra o poder dos dois métodos em determinar os efeitos aditivos. Observa-se que o Método II é mais poderoso, ou seja, estima efeitos aditivos mais próximos aos verdadeiros. O poder para os efeitos dominantes é apresentado na Figura 36, Apêndice A. O Método II é mais poderoso e ambos apresentaram diferenças médias baixas de efeito dominante.

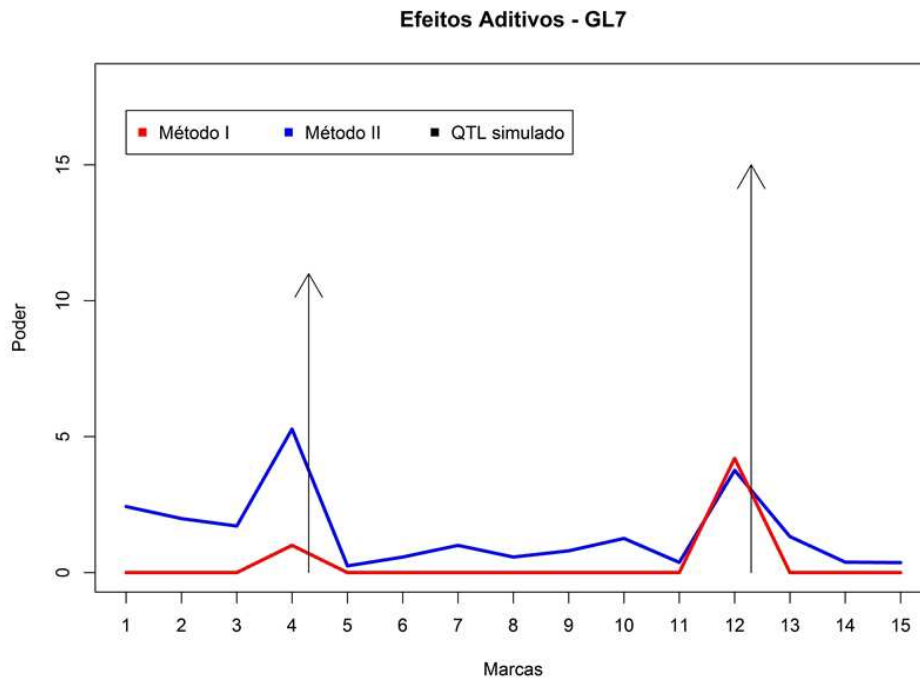


Figura 9 Poder dos Métodos I e II para determinar efeito aditivo dos QTLs, para o sétimo grupo de ligação, com 20% de perda de marcadores.

Na Figura 10 apresentam-se as posições dos QTLs estimados pelas marcas, as frequências relativas com seus respectivos intervalos de confiança exatos para proporção com 5% de significância e as diferenças de efeitos aditivos (as diferenças de efeitos dominantes encontram-se na Figura 46, Apêndice B), que ilustram os fatos descritos na discussão da Figura 8, para as marcas quatro e 12. Note que ambas as marcas estão muito próximas aos QTLs estimados. Pelo Método I, a marca 12 tem frequência de detecção alta, mas a marca quatro tem frequência baixa, não sendo significativa, mesmo com distância pequena do QTL simulado. Pelo Método II, apesar de as distâncias entre os QTL estimados e os simulados serem maiores que no Método I, as frequências de detecção das duas marcas foram altas, mostrando melhor identificação dos dois QTLs.

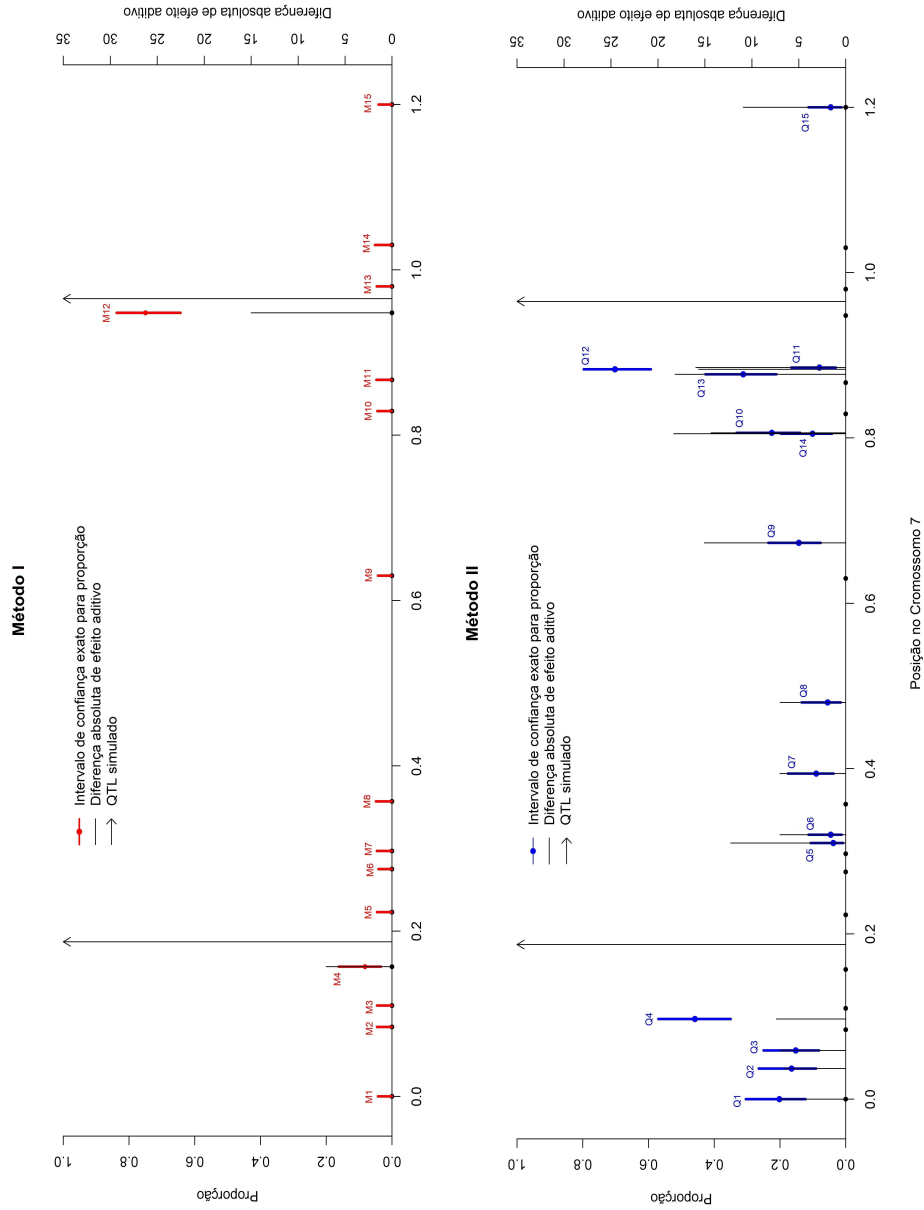


Figura 10 QTLs estimados pelos Métodos I e II com as respectivas posições (em unidade Morgan) no cromossomo sete, frequências relativas com intervalos de confiança exato para proporção com 5% de significância e diferenças absolutas dos efeitos aditivos entre os QTLs estimados e os QTLs simulados, para perda de 20% das marcas.

Os mesmos procedimentos foram feitos para o oitavo grupo de ligação. A Figura 11 mostra o poder de ambos os métodos em detectar QTLs, lembrando que o Método I é representado pela cor vermelha e o Método II pela cor azul. Observa-se que o Método I apresentou maior poder de detecção. Porém, a mesma interpretação errônea, apresentada no sétimo grupo de ligação, acontece aqui. O QTL simulado está extremamente próximo à marca sete (distância de 0,1cM), o que influenciou muito o pico apresentado pelo Método I. Entretanto, essa marca foi significativa, neste método, apenas cinco vezes em 80 seleções com essa distância pequena.

Pelo Método II, essa marca identificou QTL com distância média de 7,7cM do verdadeiro, bem mais alta que a do Método I, sendo menos enfatizada na Figura 11, mas foi significativa 39 vezes em 80 seleções. Isso mostra que o método consegue identificar o QTL verdadeiro, mesmo que um pouco distante da posição real. Essa distância média foi influenciada pelas diferentes distâncias estimadas nas 100 repetições, que variaram de $1,97e^{-11}$ cM a 33cM. Vale ressaltar que em 20 das 39 vezes em que a marca foi selecionada como significativa pelo Método II, ela identificou QTL com distância $1,97e^{-11}$ cM do verdadeiro, ou seja, identificou exatamente a posição do QTL simulado, sendo mais preciso.

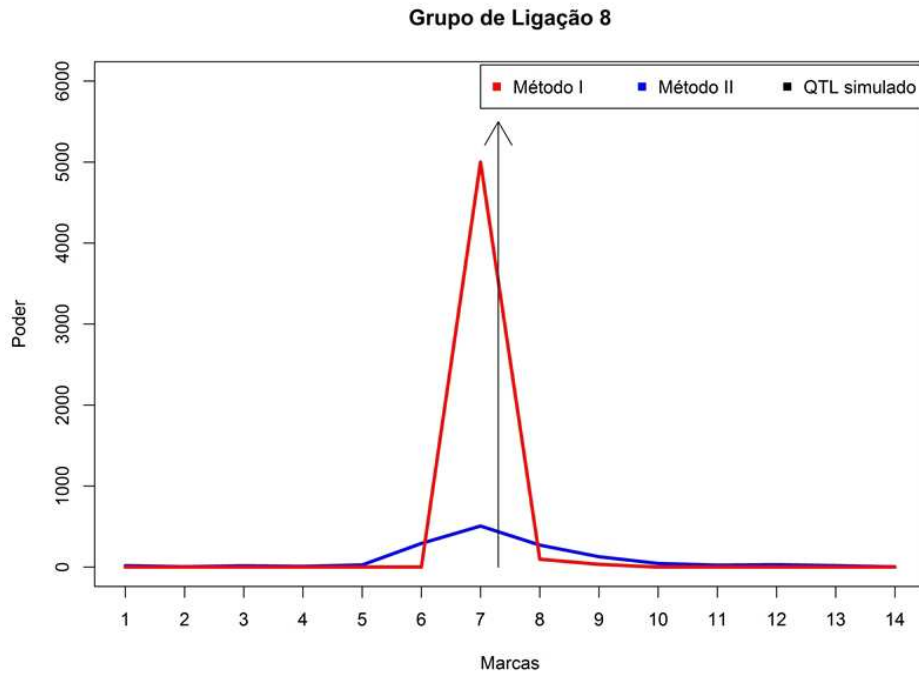


Figura 11 Poder dos Métodos I e II para detectar QTL, para o oitavo grupo de ligação simulado, com perda de 20% de marcadores.

Na Figura 12 mostra-se o poder dos dois métodos em determinar os efeitos aditivos. O Método II é mais poderoso que o Método I em determinar efeitos aditivos e, também, em determinar efeitos dominantes (Figura 37, Apêndice A). Ambos os métodos identificaram diferenças pequenas entre os efeitos aditivos do QTL estimado e do QTL simulado (Figura 13) e diferenças pequenas entre os efeitos dominantes (Figura 47, Apêndice B).

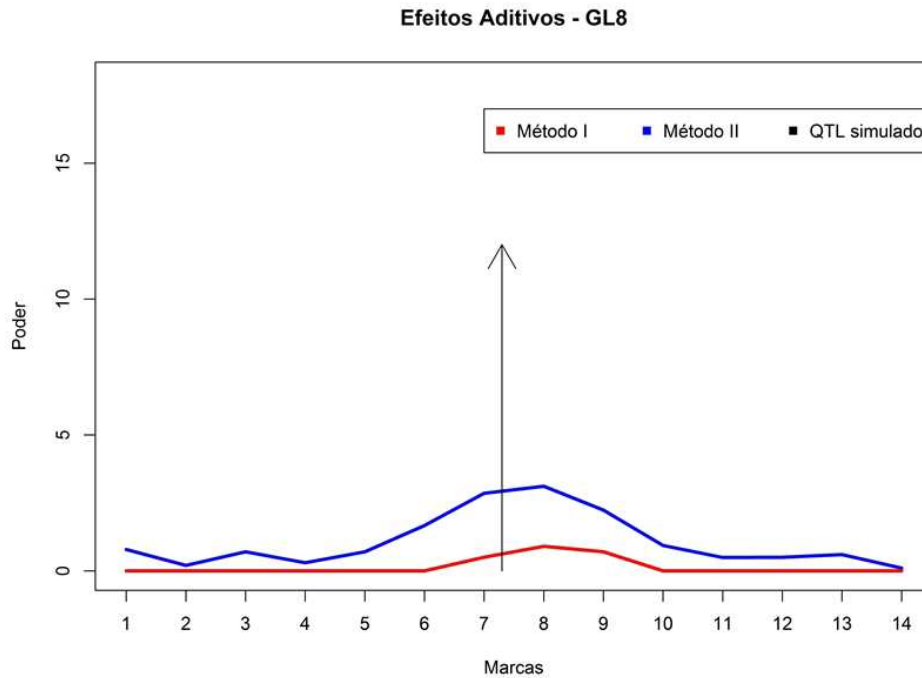


Figura 12 Poder dos Métodos I e II para detectar QTL, para o oitavo grupo de ligação simulado, com perda de 20% de marcadores.

Para mostrar o problema apresentado acima com a marca sete, criou-se a Figura 13 com as posições, no cromossomo, dos QTLs estimados por ambos os métodos, as frequências relativas com seus respectivos intervalos de confiança exatos para proporção com 5% de significância e as diferenças médias de efeitos aditivos. Um gráfico semelhante ao da Figura 13 para as diferenças de efeitos dominantes está representado na Figura 47, Apêndice B. Observe que a marca sete está muito próxima ao QTL simulado, mas possui frequência muito baixa pelo Método I, indicando que a marca não foi considerada notável; já pelo Método II, o QTL estimado está mais afastado do QTL simulado, mas foi altamente identificado pela marca, tornando-a uma marca notável.

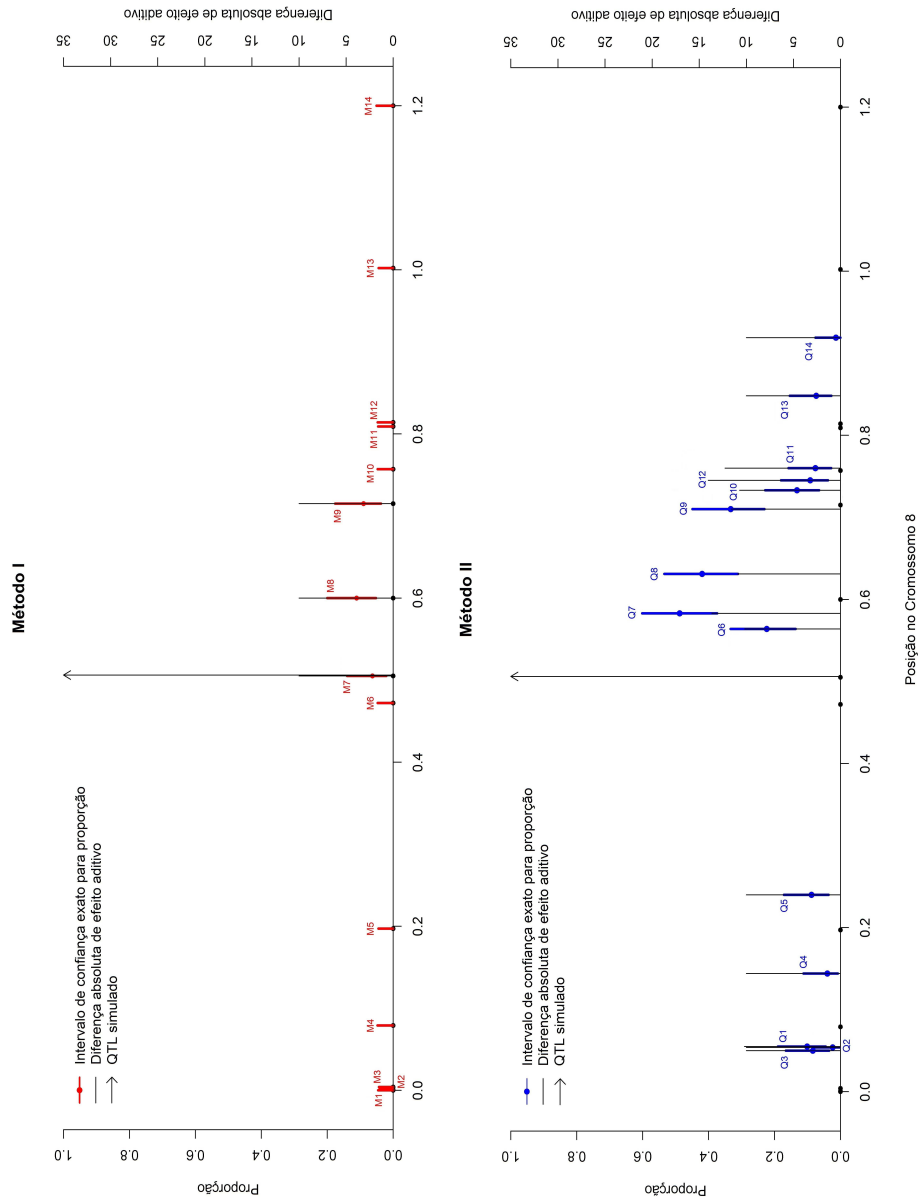


Figura 13 QTLs estimados pelos Métodos I e II com as respectivas posições (em unidade Morgan) no cromossomo oito, frequências relativas com intervalos de confiança exato para proporção com 5% de significância e diferenças absolutas dos efeitos aditivos entre os QTLs estimados e os QTLs simulados, para perda de 20% das marcas.

Na Figura 14 mostra-se o poder de detecção de ambos os métodos para o nono grupo de ligação. O Método I teve maior poder de detecção que o Método II. A marca dez é considerada uma marca notável, com frequência de detecção de 51 em 74 seleções e distância de 5cM do QTL verdadeiro, pelo Método I, e com frequência de 44 em 74 seleções e distância de 9,6cM do QTL verdadeiro, pelo Método II.

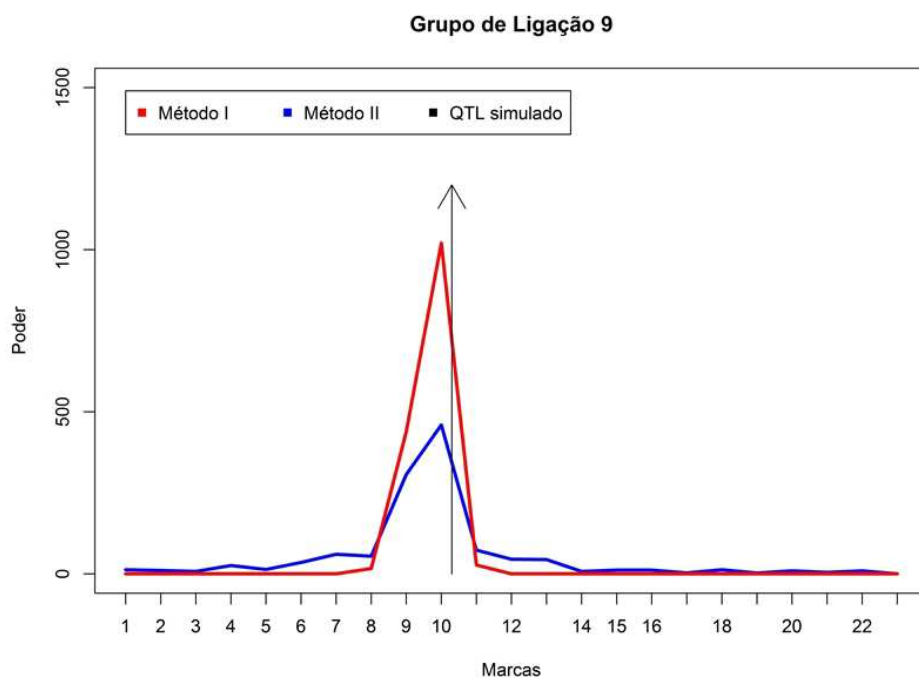


Figura 14 Poder dos Métodos I e II para detectar QTL, para o nono grupo de ligação simulado, com 20% de perda de marcadores.

O poder dos dois métodos em determinar os efeitos aditivos é apresentado na Figura 15. Ambos apresentam poderes similares, mas diferenças médias altas de efeitos aditivos (Figura 16). Para os efeitos dominantes, o poder é apresentado na Figura 38, Apêndice A, para ambos os métodos, sendo considerados similares e com diferenças médias baixas.

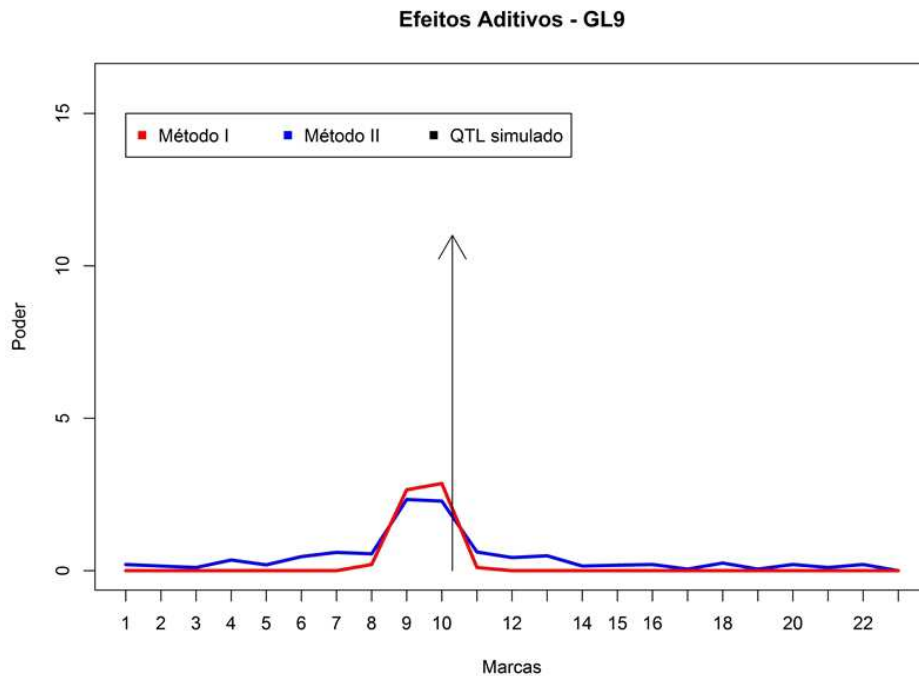


Figura 15 Poder dos Métodos I e II para determinar efeito aditivo dos QTLs, para o nono grupo de ligação, com 20% de perda de marcadores.

Na Figura 16 resumem-se as características de identificação de QTLs de ambos os métodos: as posições dos QTLs estimados, as frequências relativas com seus respectivos intervalos de confiança exatos para proporção com 5% de significância e as diferenças médias de efeitos aditivos (um gráfico similar foi feito para as diferenças médias de efeitos dominantes, apresentado na Figura 48, Apêndice B). Note que as marcas nove e dez identificaram QTLs com maior frequência em ambos os métodos e a marca dez identificou QTL mais próximo do verdadeiro. As diferenças de efeitos aditivos foram altas e as de efeitos dominantes foram baixas, nos dois métodos.

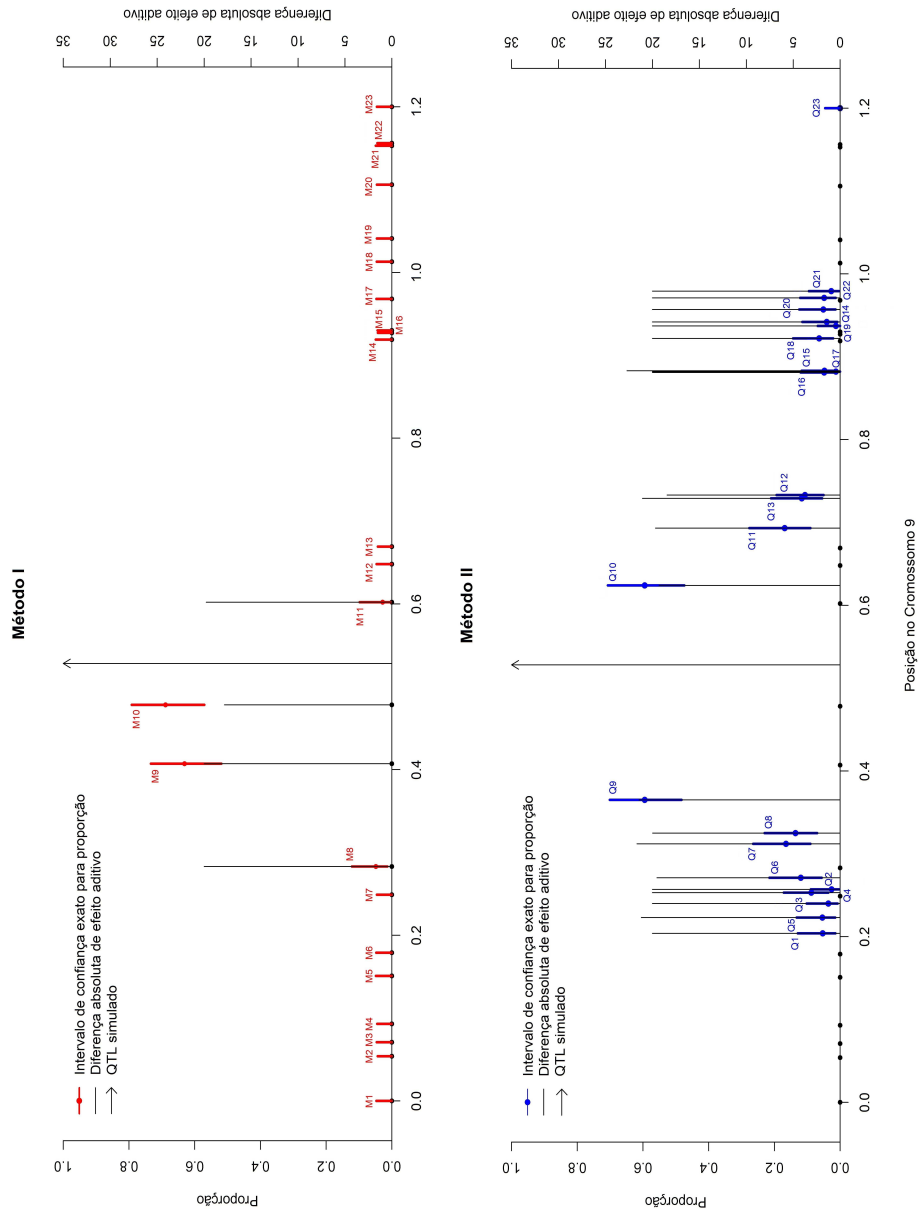


Figura 16 QTLs estimados pelos Métodos I e II com as respectivas posições (em unidade Morgan) no cromossomo nove, frequências relativas com intervalos de confiança exato para proporção com 5% de significância e diferenças absolutas dos efeitos aditivos entre os QTLs estimados e os QTLs simulados, para perda de 20% das marcas.

4.1.2 Perda de 80% das marcas

Retirando-se 80% das marcas em cada grupo de ligação, restaram 81 marcas no genoma. Avaliaram-se a distância, a diferença absoluta entre os efeitos aditivos e a diferença absoluta entre os efeitos dominantes do QTL simulado com o QTL estimado (ou com a marca selecionada), para os dois métodos.

Na Figura 17 mostra-se o poder para detectar QTL no primeiro grupo de ligação (GL1), para o Método I (em vermelho) e para o Método II (em azul). Quanto maior o pico, menor a distância entre eles e maior a frequência da marca; assim, se o QTL encontrado for exatamente o verdadeiro, a distância será zero e o pico tenderá ao infinito. Nesta figura, observa-se que o Método II encontrou os dois QTLs simulados com maior poder, enquanto o Método I encontrou os dois QTLs, mas com poder baixo para o primeiro.

De fato, o marcador dez foi significativo 14 vezes em 16 seleções, pelo Método I, mas com distância de 11,7cM do verdadeiro QTL, o que penalizou o pico no gráfico. No Método II este marcador encontrou um QTL 16 vezes em 16 seleções, com distância de 1,7cM do verdadeiro, praticamente identificando o verdadeiro. Vale ressaltar que em 12 das 16 vezes em que a marca 10 foi significativa, ela identificou QTL com distância de $3,02e^{-11}$ cM do verdadeiro. O marcador 12 é considerado um provável QTL (16 vezes significativo em 17 seleções), pelo Método I, com distância média de 4,7cM do verdadeiro QTL; no Método II, esse marcador encontrou um QTL com distância média do verdadeiro de 0,86cM, sendo significativo 14 vezes em 17 seleções (Tabela 3, Apêndice C). Vale ressaltar que em oito das 14 vezes em que a marca 12 foi significativa, ela identificou QTL com distância de $4,13e^{-11}$ cM do verdadeiro.

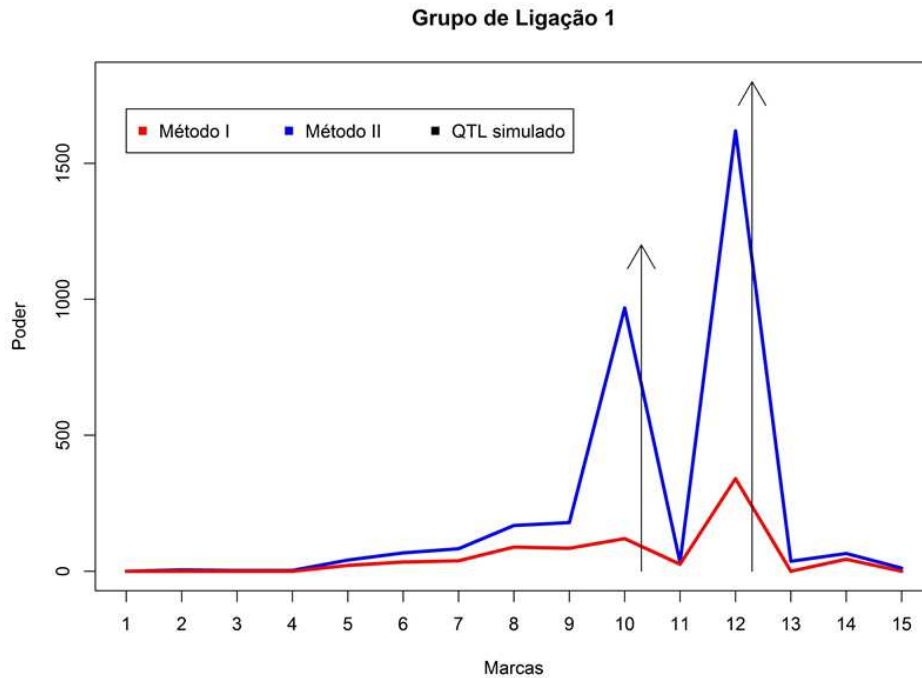


Figura 17 Poder dos Métodos I e II para detectar QTL, para o primeiro grupo de ligação simulado, com 80% de perda de marcadores.

Na Figura 18 mostra-se o poder dos métodos em determinar efeito aditivo e os métodos são parecidos. Para os efeitos dominantes, o poder é apresentado na Figura 39, Apêndice A, sendo também considerados similares.

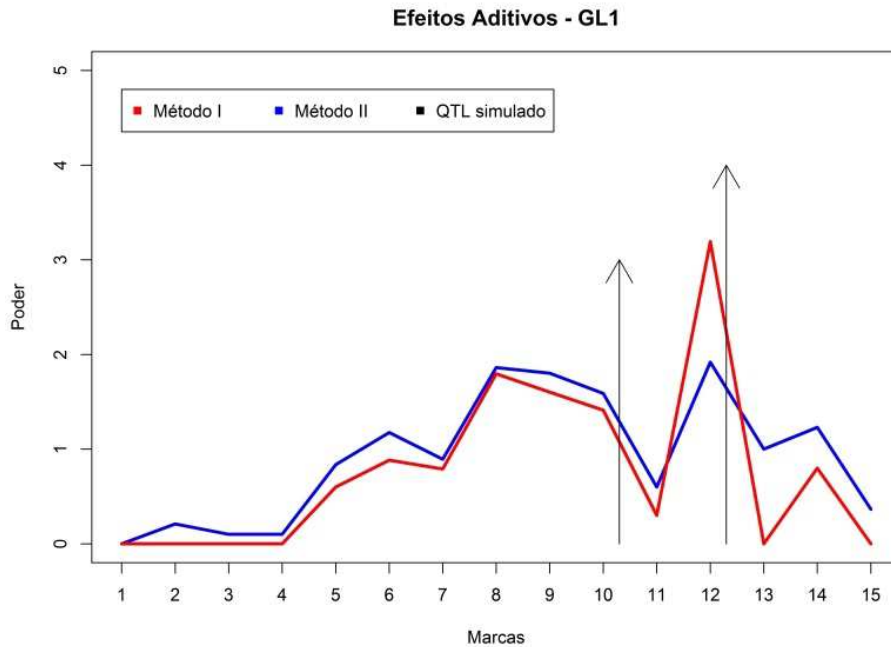


Figura 18 Poder dos Métodos I e II para determinar efeito aditivo dos QTLs, para o primeiro grupo de ligação, com 80% de perda de marcadores.

As características relacionadas aos QTLs estimados por ambos os métodos encontram-se na Figura 19, onde estão representadas as posições dos QTLs estimados pelos métodos, as frequências relativas com seus respectivos intervalos de confiança exatos para proporção com 5% de significância e as diferenças médias de efeitos aditivos. Observe que as marcas dez e 12 estimaram QTLs em ambos os métodos, sendo mais bem estimados no Método II, já que as distâncias com relação ao QTL simulado foram menores. Note que os intervalos de confiança são maiores em decorrência de o número de seleções ser baixo, ou seja, menores tamanhos amostrais na simulação. Apenas a marca um não identificou QTL no Método II por se encontrar muito distante do QTL simulado. Um gráfico semelhante com as diferenças médias de efeitos dominantes encontra-se na Figura 49, Apêndice B, mostrando que as diferenças médias foram altas em ambos os métodos.

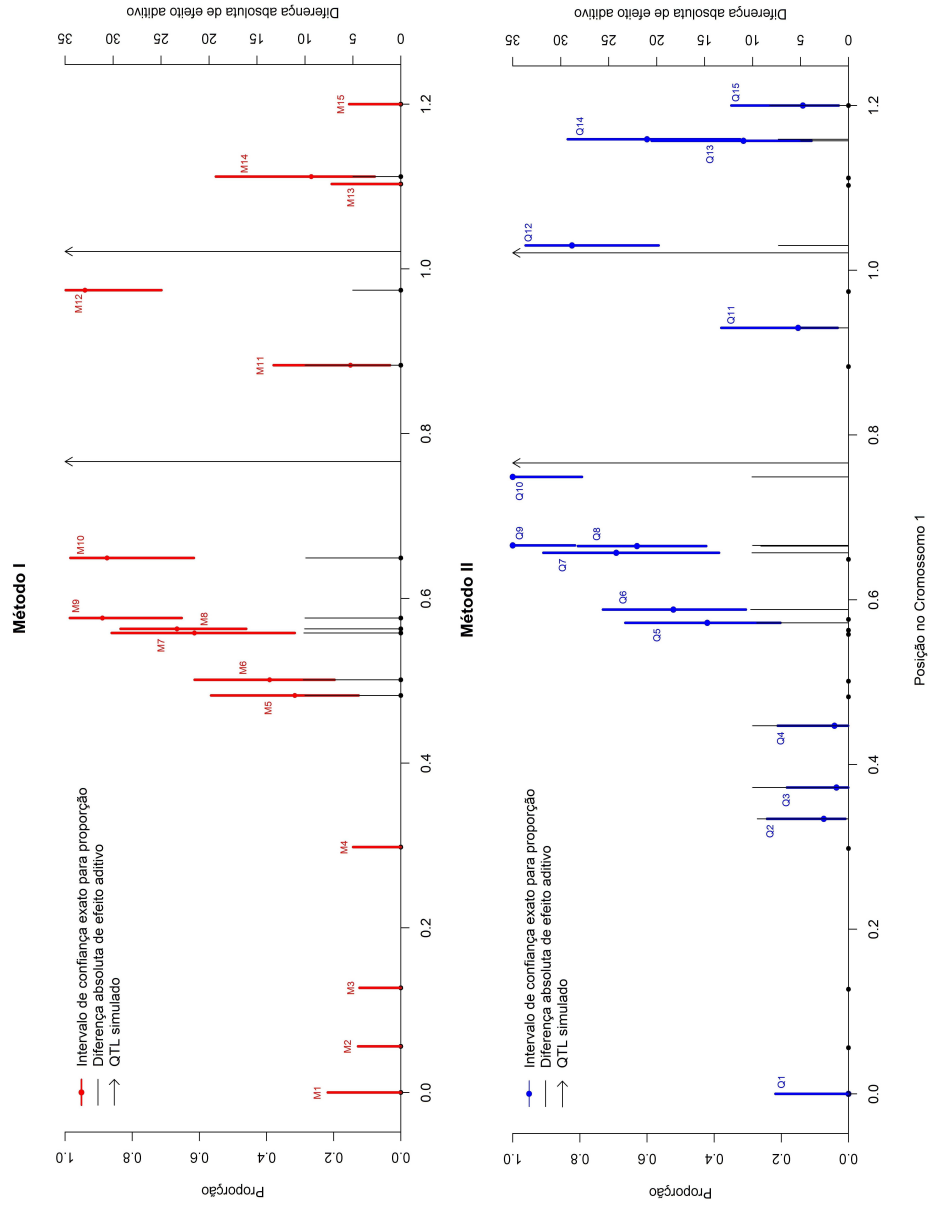


Figura 19 QTLs estimados pelos Métodos I e II com as respectivas posições (em unidade Morgan) no cromossomo um, frequências relativas com intervalos de confiança exato para proporção com 5% de significância e diferenças absolutas dos efeitos aditivos entre os QTLs estimados e os QTLs simulados, para perda de 80% das marcas.

Na Figura 20 mostra-se o poder de detecção de QTL para os dois métodos no quinto grupo de ligação. O Método II foi mais poderoso em detectar QTL que o Método I. A frequência de detecção da marca dois pelo Método I foi de oito em 15 seleções com distância de 12,8cM do QTL verdadeiro, enquanto a frequência pelo Método II foi de 11 em 15 seleções com distância média de 6,5cM do verdadeiro (Tabela 3, Apêndice C; Figura 22). Vale ressaltar que em quatro das 11 vezes em que foi significativa, a marca detectou um QTL com distância de $2,71e^{-11}$ cM do verdadeiro.

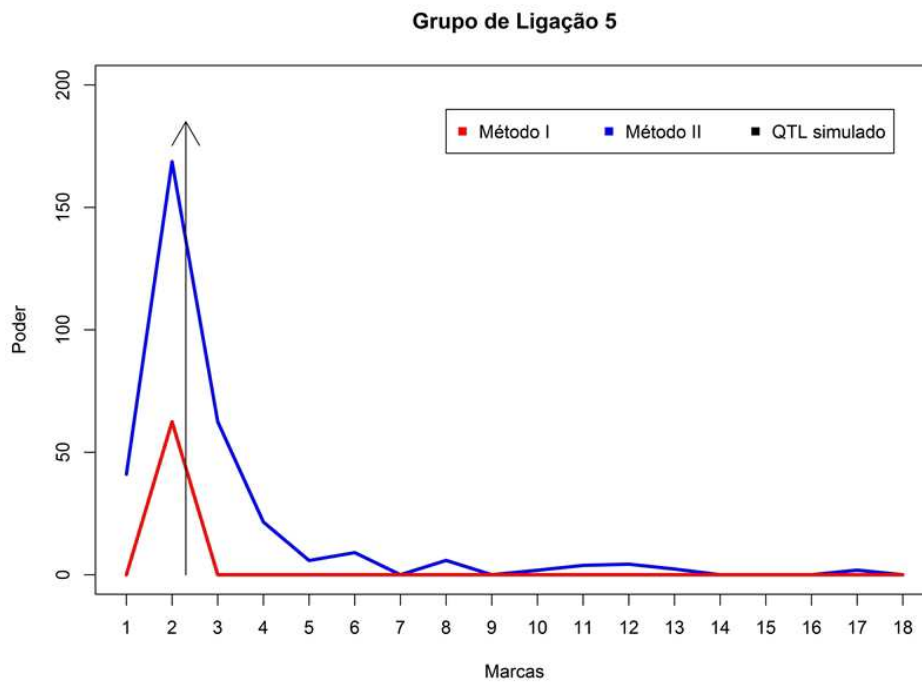


Figura 20 Poder dos Métodos I e II para detectar QTL, para o quinto grupo de ligação simulado, com 80% de perda de marcadores.

O poder dos dois métodos em determinar efeitos aditivos é representado na Figura 21. O Método II é considerado mais poderoso que o Método I. O poder em determinar os efeitos dominantes é apresentado na Figura 40, Apêndice A,

sendo o Método II considerado mais poderoso que o Método I.

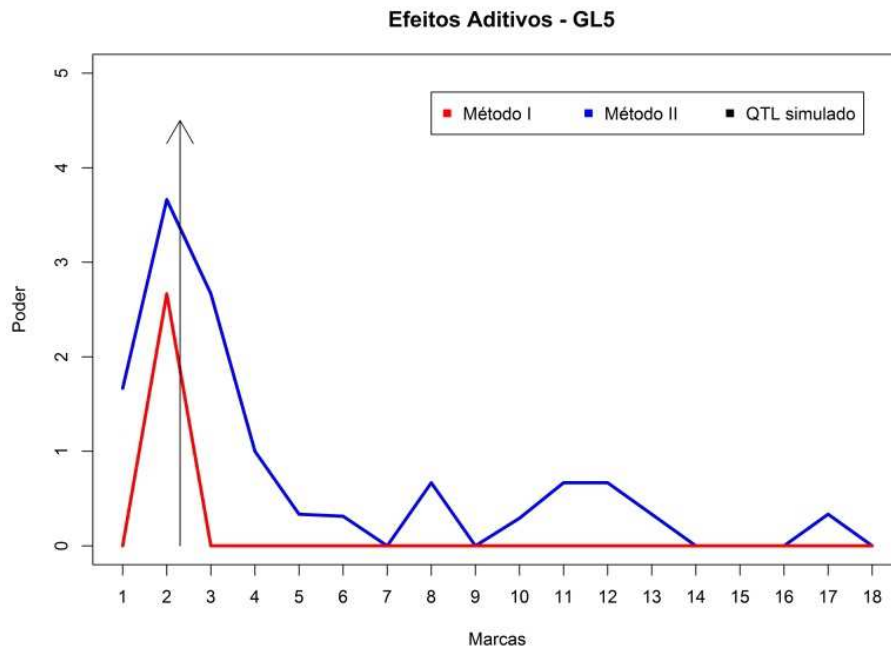


Figura 21 Poder dos Métodos I e II para determinar efeito aditivo dos QTLs, para o quinto grupo de ligação, com 80% de perda de marcadores.

As posições dos QTLs estimados por ambos os métodos, as frequências de detecção com seus respectivos intervalos de confiança exatos para proporção com 5% de significância e as diferenças de efeitos aditivos estão representados na Figura 22. Observe que a marca dois, pelo Método II, estimou um QTL mais próximo do verdadeiro que o Método I. As diferenças médias de efeitos aditivos foram baixas para ambos os métodos. Um gráfico semelhante para as diferenças de efeitos dominantes é apresentado na Figura 50, Apêndice B. As diferenças médias foram semelhantes para ambos os métodos, sendo consideradas baixas.

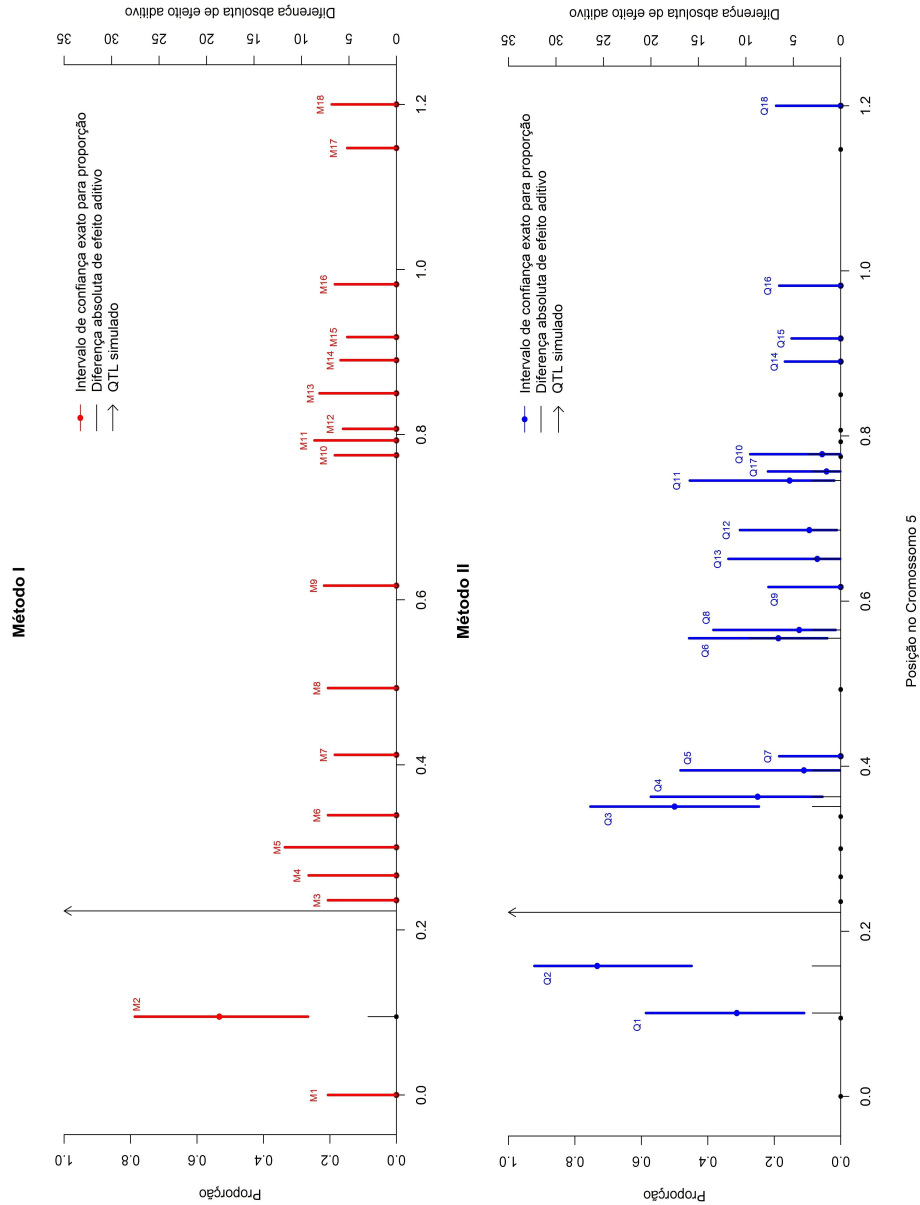


Figura 22 QTLs estimados pelos Métodos I e II com as respectivas posições (em unidade Morgan) no cromossomo cinco, frequências relativas com intervalos de confiança exato para proporção com 5% de significância e diferenças absolutas dos efeitos aditivos entre os QTLs estimados e os QTLs simulados, para perda de 80% das marcas.

Na Figura 23 representa-se o poder de detecção de QTL pelos dois métodos no sétimo grupo de ligação. O Método II identificou melhor os QTLs simulados, embora a figura tenha apresentado melhor poder de detecção para o Método I na marca 12. Primeiro, pelo Método I não se identificou o primeiro QTL; segundo, o pico na marca 12 não está bem representado, já que ela foi significativa apenas seis vezes em 19 seleções. Isso se deve ao fato de que a marca 12 está muito próxima ao QTL simulado (1,7cM de distância), influenciando o pico. Pelo Método II, essa marca foi significativa 13 vezes em 19 seleções, mas com distância média de 9cM do verdadeiro QTL, o que fez a altura do pico ser mais baixa. Vale ressaltar que em seis das 13 vezes em que a marca 12 foi significativa, ela identificou QTL com distância de $6,86e^{-11}$ cM do verdadeiro.

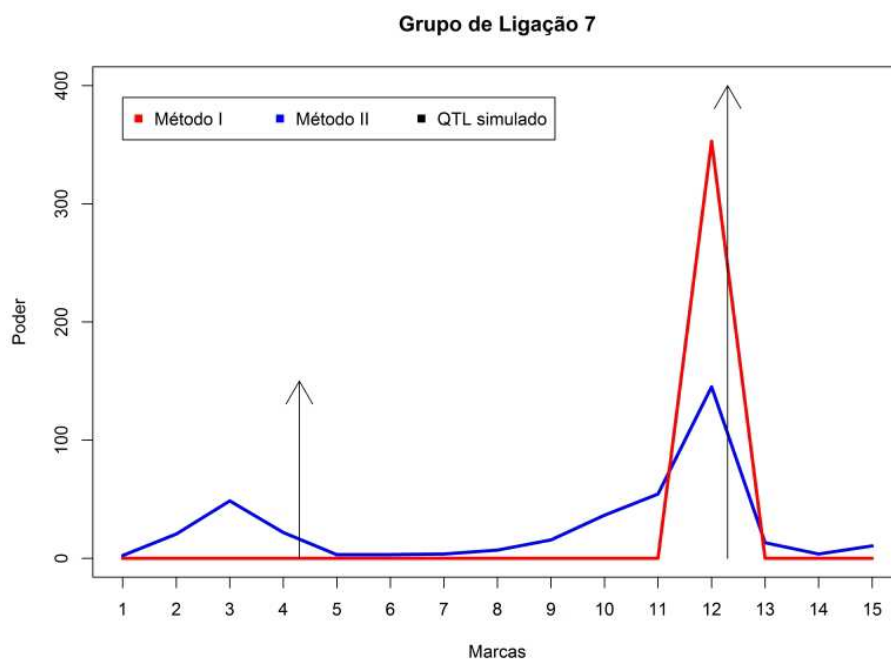


Figura 23 Poder dos Métodos I e II para detectar QTL, para o sétimo grupo de ligação simulado, com 80% de perda de marcadores.

Pela Figura 24 mostra-se o poder dos métodos em determinar efeitos adi-

tivos no sétimo grupo de ligação. O Método II foi ligeiramente melhor que o Método I. O gráfico de poder para efeitos dominantes encontra-se na Figura 41, Apêndice A.

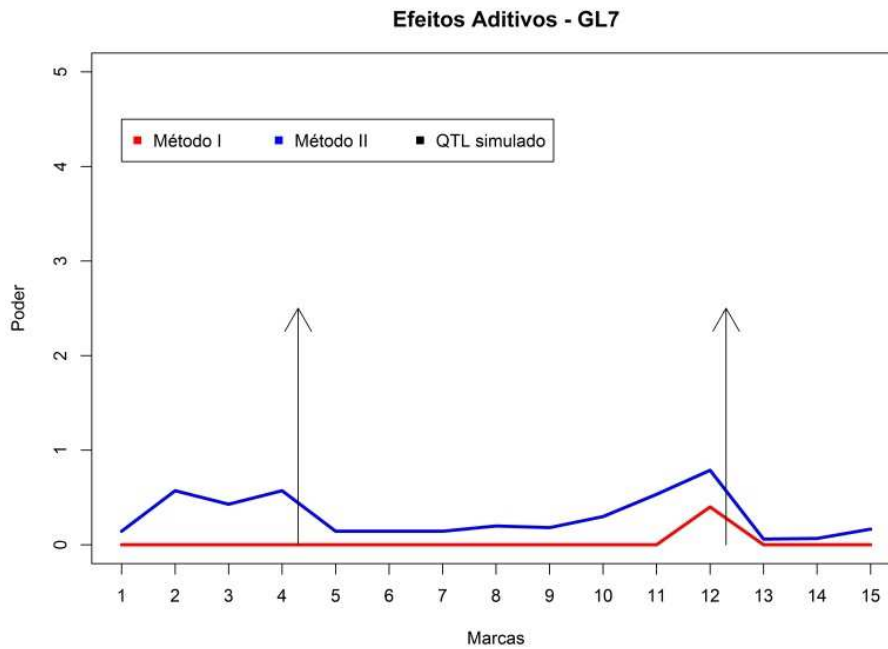


Figura 24 Poder dos Métodos I e II para determinar efeito aditivo dos QTLs, para o sétimo grupo de ligação, com 80% de perda de marcadores.

Na Figura 25 é representado o resumo das características dos QTLs estimados: suas posições no cromossomo, as frequências relativas de detecção com os respectivos intervalos de confiança exatos para proporção com 5% de significância e as diferenças médias dos efeitos aditivos do QTL estimado com o QTL simulado. Um gráfico similar para as diferenças médias dos efeitos dominantes é representado na Figura 51, Apêndice B. Observa-se que a marca 12 encontra-se muito próxima ao QTL simulado, mas foi pouco identificada pelo Método I. Já pelo Método II, o QTL estimado encontra-se mais distante, mas com frequência maior. Os efeitos aditivos foram altos para ambos os métodos.

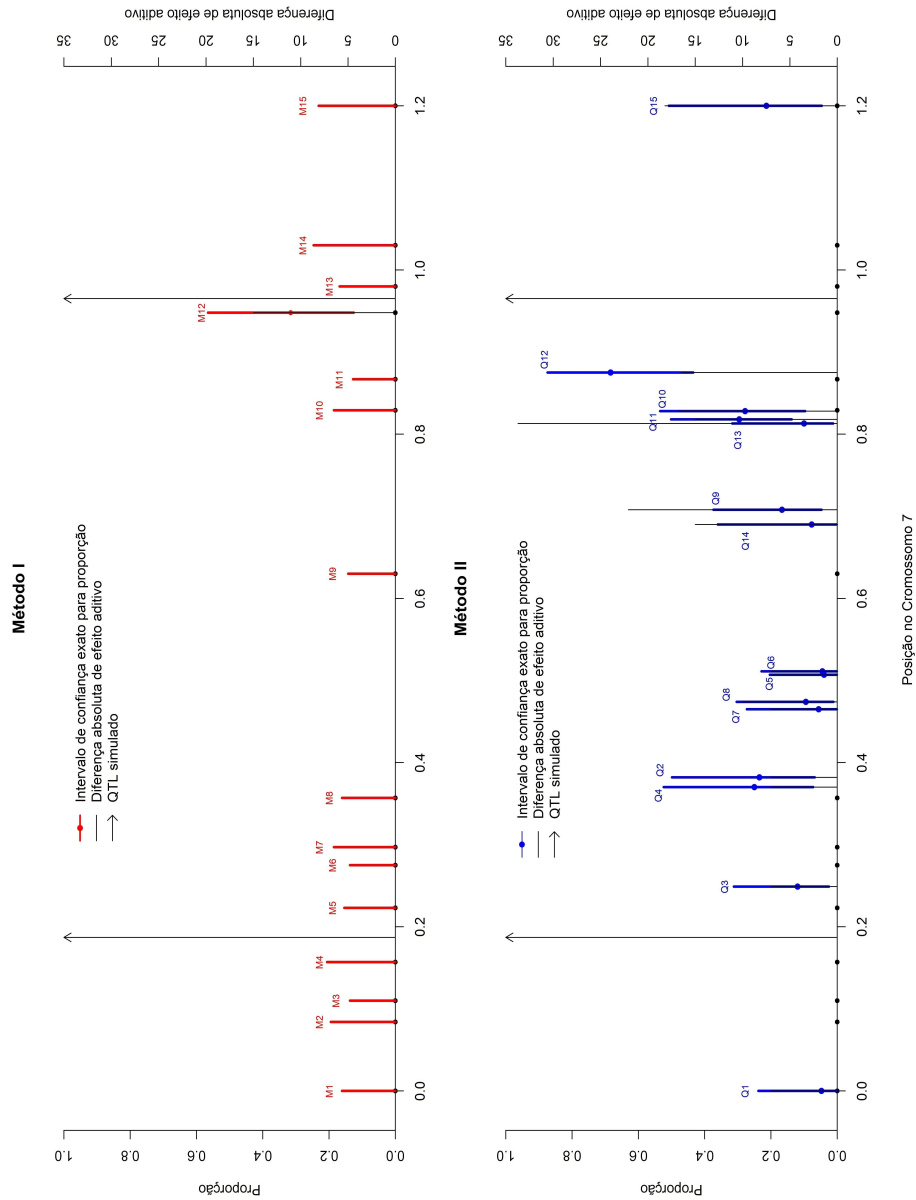


Figura 25 QTLs estimados pelos Métodos I e II com as respectivas posições (em unidade Morgan) no cromossomo sete, frequências relativas com intervalos de confiança exato para proporção com 5% de significância e diferenças absolutas dos efeitos aditivos entre os QTLs estimados e os QTLs simulados, para perda de 80% das marcas.

Os mesmos procedimentos foram feitos para o oitavo grupo de ligação. Na Figura 26 representa-se o poder de detecção dos dois métodos e o Método II é considerado mais poderoso que o Método I. Pelo Método I identificou-se o QTL na marca nove, mas com baixa frequência (três em 12 seleções) e com distância de 20,9cM do QTL verdadeiro (Figura 28). Pelo Método II, a marca sete identificou QTL com frequência de oito em 19 seleções e distância média de 12,3cM do verdadeiro (Tabela 3, Apêndice C). Vale ressaltar que em três das 8 vezes em que a marca foi significativa, ela identificou QTL com distância de $1,97e^{-11}$ cM do verdadeiro.

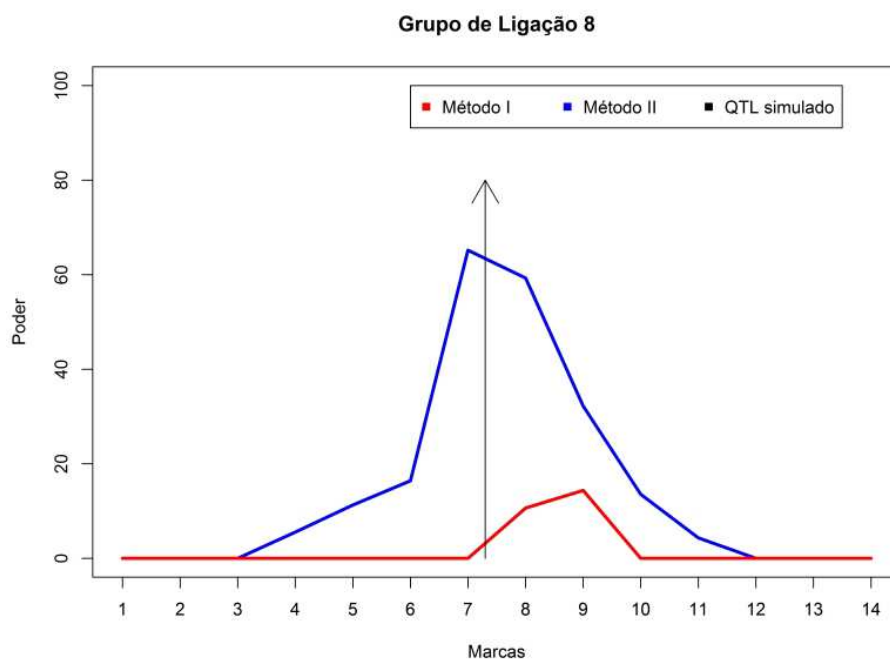


Figura 26 Poder dos Métodos I e II para detectar QTL, para o oitavo grupo de ligação simulado, com 80% de perda de marcadores.

O poder em determinar os efeitos aditivos, pelos dois métodos, é representado na Figura 27. O Método II pode ser considerado mais poderoso que o Método I. O mesmo procedimento foi feito para os efeitos dominantes e o poder está re-

presentado na Figura 42, Apêndice A, e o Método II é considerado mais poderoso que o Método I.

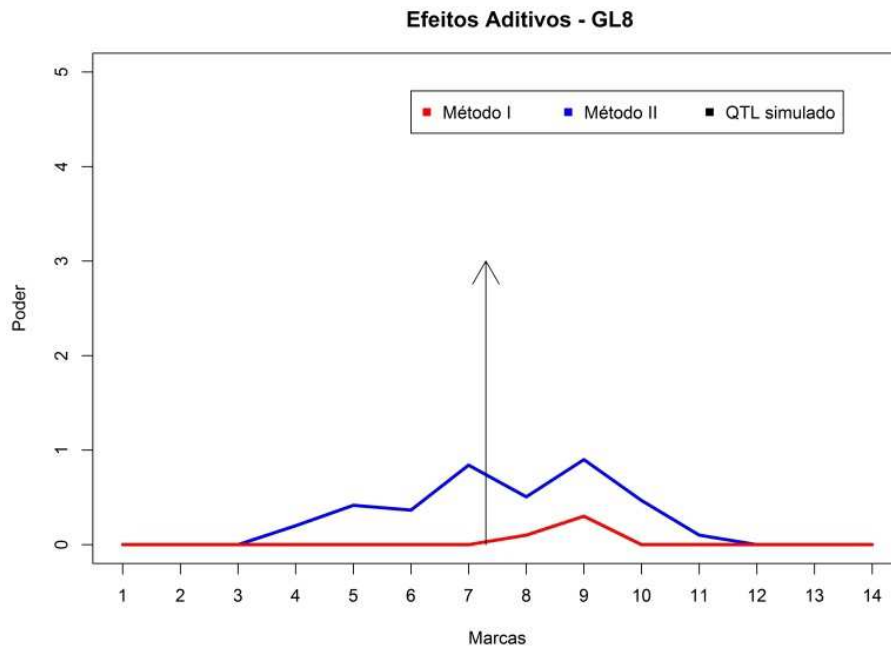


Figura 27 Poder dos Métodos I e II para determinar efeito aditivo dos QTLs, para o oitavo grupo de ligação, com 80% de perda de marcadores.

As principais características dos QTLs estimados estão representadas na Figura 28: suas posições no cromossomo, as frequências relativas com seus intervalos de confiança exatos para proporção com 5% de significância e as diferenças médias dos efeitos aditivos. Um gráfico semelhante é feito para as diferenças médias dos efeitos dominantes e representado na Figura 52, Apêndice B. Observe que o QTL estimado pela marca oito está mais próximo do QTL verdadeiro. Ele foi estimado cinco vezes em nove seleções com uma distância de 8,4cM. Isso leva a pensar que se esta marca tivesse sido selecionada mais vezes, ela seria notável, já que foi muito significativa dentre as seleções.

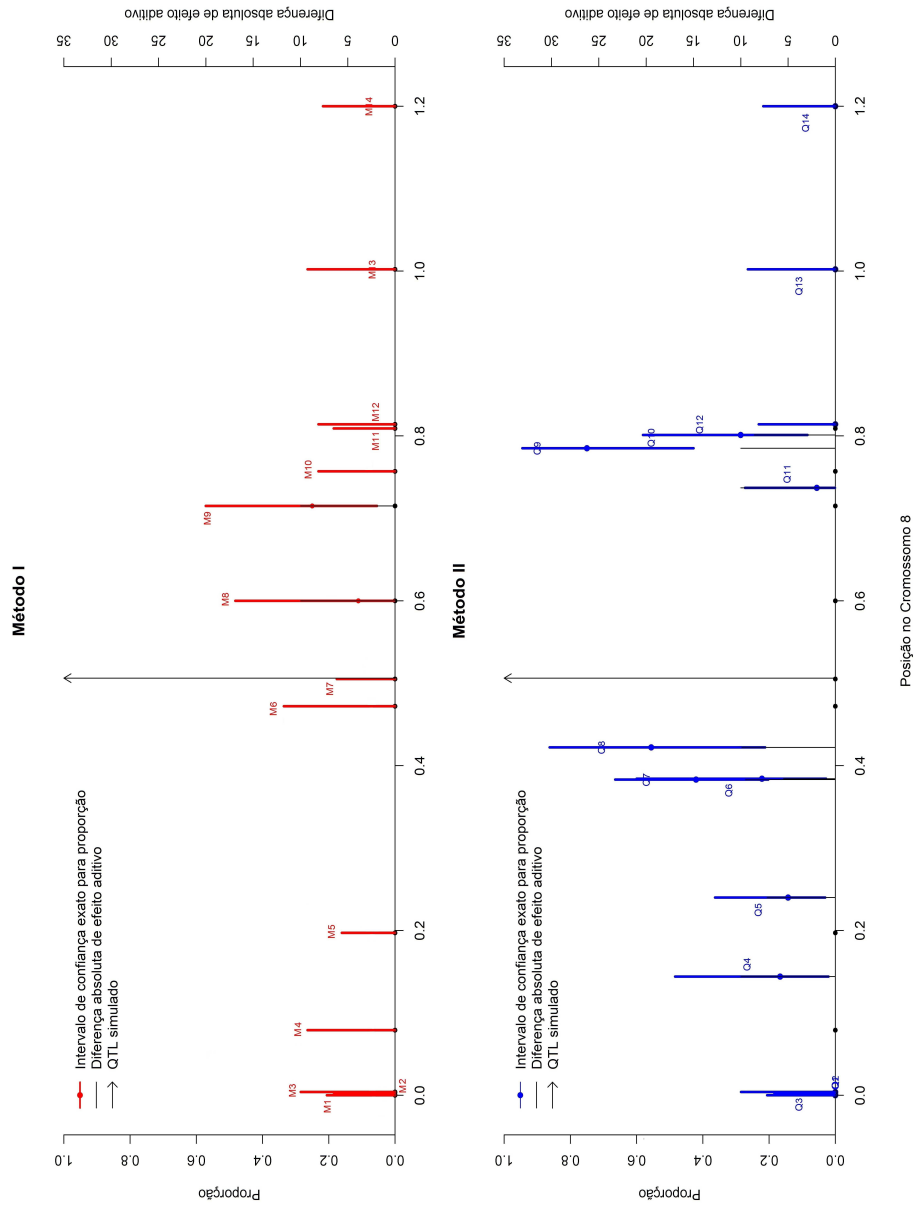


Figura 28 QTLs estimados pelos Métodos I e II com as respectivas posições (em unidade Morgan) no cromossomo oito, frequências relativas com intervalos de confiança exato para proporção com 5% de significância e diferenças absolutas dos efeitos aditivos entre os QTLs estimados e os QTLs simulados, para perda de 80% das marcas.

Na Figura 29 é representado o poder de detecção de QTL dos métodos no nono grupo de ligação. Observa-se que o Método I foi mais poderoso que o Método II e que as marcas sete, oito, nove, dez, 11 e 12 parecem identificar o QTL simulado, já que foram bem significativas dentro das seleções, apesar das distâncias maiores (Tabela 3, Apêndice C).

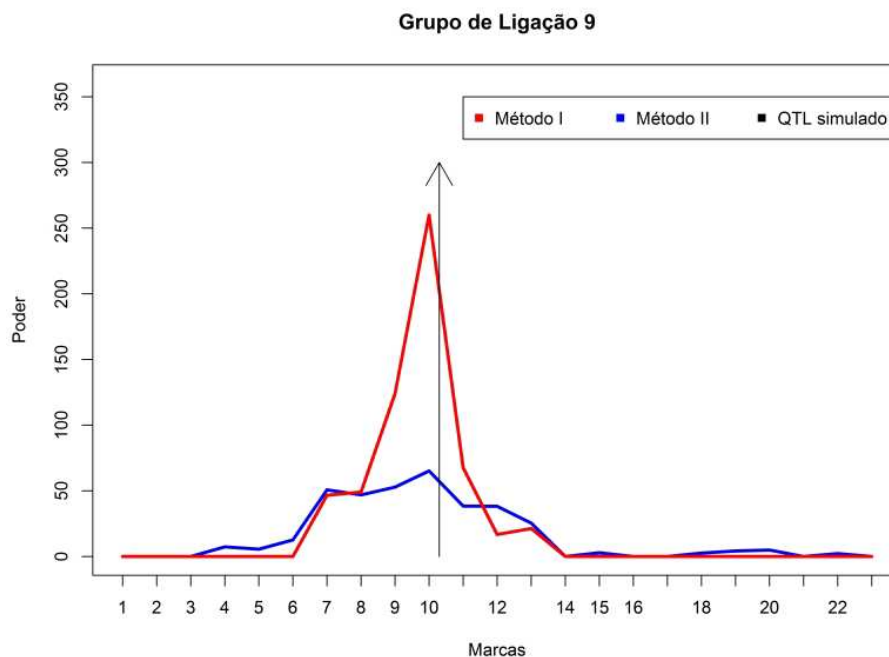


Figura 29 Poder dos Métodos I e II para detectar QTL, para o nono grupo de ligação simulado, com 80% de perda de marcadores.

O gráfico do poder em determinar efeitos aditivos, para os dois métodos, está representado na Figura 30 abaixo, assim como o poder em determinar efeitos dominantes está representado na Figura 43, Apêndice A. Os métodos são similares em determinar efeitos aditivos, apesar de as diferenças médias serem altas, e são similares em determinar efeitos dominantes com diferenças médias baixas.

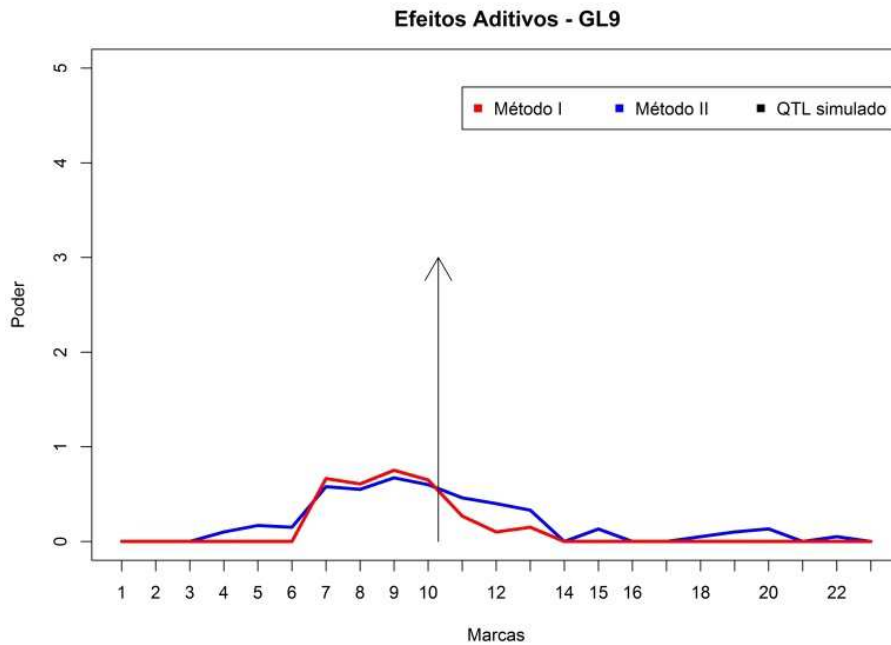


Figura 30 Poder dos Métodos I e II para determinar efeito aditivo dos QTLs, para o nono grupo de ligação, com 80% de perda de marcadores.

As características dos QTLs estimados (posições, frequências, diferenças de efeitos aditivos) estão representadas na Figura 31 e as frequências possuem intervalos de confiança exatos para proporção com 5% de significância. Um gráfico similar para as diferenças de efeitos dominantes é mostrado na Figura 53, Apêndice B. Observe que a marca dez está próxima ao QTL verdadeiro e com frequência alta, o que a torna notável, pelo Método I. As marcas sete, oito e nove, também, tiveram frequências altas, porém estão mais afastadas do QTL. Pelo Método II, as marcas sete a 12 tiveram frequências altas, porém estão afastadas do verdadeiro QTL, levando a pensar que conseguiram identificar o QTL, mas não localizá-lo perfeitamente. As diferenças de efeitos aditivos foram altas em ambos os métodos.

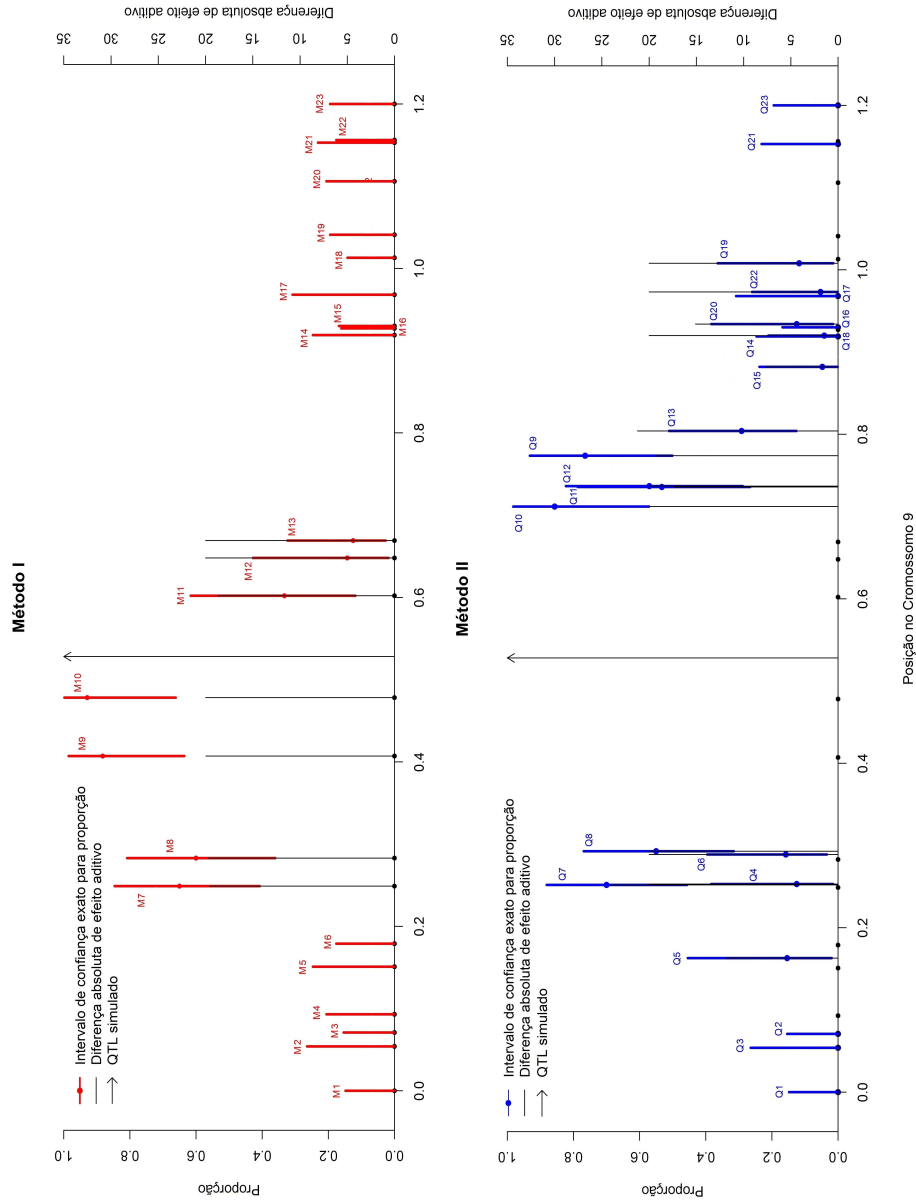


Figura 31 QTLs estimados pelos Métodos I e II com as respectivas posições (em unidade Morgan) no cromossomo nove, frequências relativas com intervalos de confiança exato para proporção com 5% de significância e diferenças absolutas dos efeitos aditivos entre os QTLs estimados e os QTLs simulados, para perda de 80% das marcas.

4.2 Dados reais

Como mostrado em Lara et al. (2014), dos 59 marcadores usados, 17 identificaram QTLs para resistência ao mofo-branco, utilizando o Método II, de acordo com a estatística de teste de Wald (Figura 33). Já pelo Método I, nenhuma marca foi significativa (Figura 32), já que a estatística de teste de Wald não foi maior que $\chi^2_{(0,95;2)} = 5,99$ para nenhuma das marcas.

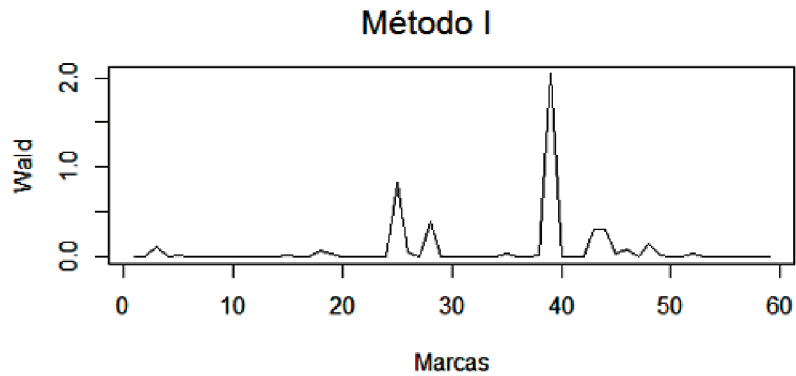


Figura 32 Poder dos Métodos I e II para detectar QTL, para o nono grupo de ligação simulado, com 80% de perda de marcadores.

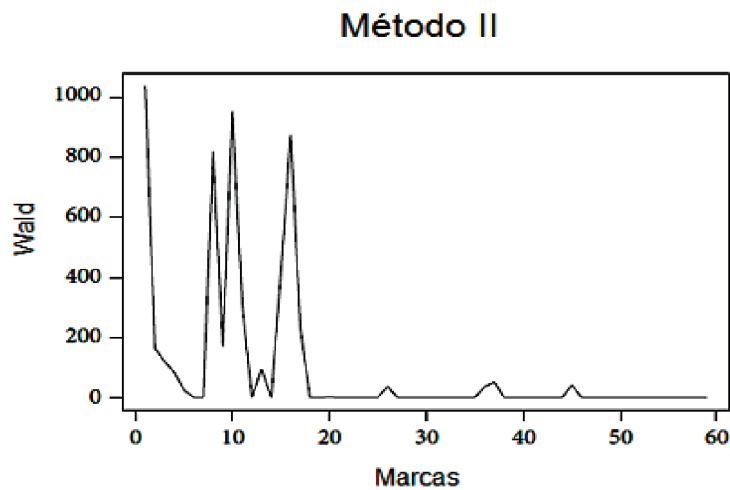


Figura 33 Poder dos Métodos I e II para detectar QTL, para o nono grupo de ligação simulado, com 80% de perda de marcadores.

5 DISCUSSÃO

No método de Xu (2003), há conhecimento a priori de que a maioria dos marcadores tem efeitos insignificantes. A inversa da variância normalmente serve como um coeficiente de penalidade. Se o marcador tem efeito notável, não será penalizado tão severamente quanto um marcador com efeito insignificante (XU, 2003).

No método de Wang et al. (2005), em cada intervalo formado por dois marcadores adjacentes, assume-se conter um QTL, analisando-os simultaneamente. Como muitos intervalos, na realidade, não contém QTLs, então eles são penalizados e encolhidos para zero pela metodologia bayesiana shrinkage.

Em Doerge, Zeng e Weir (1997), os autores utilizaram a técnica *moving away from the marker* em marcadores individuais, sendo considerado outro método de mapeamento por marcas simples, avaliando se a marca está ou não ligada a um QTL (WU; MA; CASELLA, 2007).

Baseando-se nisso, mas utilizando todas as marcas simultaneamente, foi apresentada neste trabalho uma nova técnica de identificação de QTLs, adaptando os métodos de Wang et al. (2005) e Xu (2003). A ideia é utilizar todas as marcas para identificar QTLs e, em vez de intervalos fixos criados por marcas adjacentes como em Wang et al. (2005), considera-se cada marca como pivô para a busca de QTLs à direita e à esquerda dela, em intervalos designados pela marca e por uma fração de recombinação preestabelecida. À medida que se afasta da marca (fração de recombinação aumenta), em cada posição, determinada por um caminhamento dentro destes intervalos, é analisada a presença ou não de QTL (técnica *moving away from the marker*). Se o genoma é pouco saturado, sendo marcadores adjacentes muito distantes um do outro, a fração de recombinação para definir o intervalo pode ser mais alta que a apresentada aqui, para que se garanta que todo

o genoma seja visitado na busca de QTL; se o genoma for mais saturado, pode-se utilizar uma fração de recombinação menor para que não haja tanta sobreposição de intervalos.

Nesta técnica, adota-se a abordagem bayesiana *shrinkage* que, assim como em Wang et al. (2005), força posições que não contém QTL (ou que não têm influência de QTLs próximos), dentro dos intervalos, a ter seus efeitos encolhidos para zero. No método de Xu (2003), apenas marcadores notáveis têm seus efeitos expressos nos gráficos, sendo todos os outros encolhidos a zero.

Aqui, praticamente todas as marcas encontraram algum QTL dentro do seu intervalo e apenas as marcas muito distantes do QTL verdadeiro (nos dados simulados) não identificaram QTL. Isso quer dizer que os intervalos são influenciados pela presença do QTL verdadeiro, levando a obter QTLs "fantasmas" à medida que se afasta da região do QTL. Entretanto, isso não parece ser um problema, pois a frequência com que estes QTLs foram identificados nas análises dos dados simulados foi baixa e, quando a frequência de detecção era alta, significava que se estava mais próximo da região do QTL verdadeiro. Outro fato é que, como os intervalos podem se sobrepor em virtude da distância estabelecida pela fração de recombinação escolhida, muitas marcas podem estar identificando o mesmo QTL, podendo ser, por isso, que vários QTLs foram identificados muito próximos.

De acordo com o fato citado acima, identificou-se um padrão nas análises com o método II: à medida que se aproxima do verdadeiro QTL, as marcas produzem distâncias cada vez menores entre o QTL estimado e o QTL simulado, ou seja, as marcas identificam QTLs cada vez mais próximos do QTL verdadeiro, o que não acontece para o Método I (Tabelas 2 e 3, Apêndice C). Esse padrão de distâncias é menos observado para perda de 80% das marcas, mas ainda sim o apresenta. Como neste método não é necessário o uso de mapa de ligação, se não

se conhece tal mapa, este padrão leva a pensar que, mesmo com marcas desordenadas, poder-se-ia tentar reordená-las e, assim, reconstruir os grupos de ligação.

Por meio das análises com os dados reais, verificou-se que o mapeamento de regressão por múltiplas marcas, proposto por Xu (2003), não é considerado eficiente para detectar QTLs em genomas com poucas marcas, sendo a técnica *moving away from the marker bayesiana*, proposta neste trabalho, considerada melhor em tais condições.

6 CONCLUSÃO

Diante de todo o exposto, conclui-se que a técnica de identificação de QTL proposta neste trabalho, *moving away from the marker* bayesiana, seja mais recomendada que o mapeamento por regressão bayesiano de Xu (2003) quando há poucos marcadores distribuídos no genoma, pois conseguiu estimar melhor os QTLs.

Talvez seja possível reconstruir mapas de ligação utilizando esta técnica, já que se consegue um padrão de reordenação das marcas de acordo com as distâncias encontradas entre as marcas e os prováveis QTLs.

Trabalhos futuros poderão ser feitos comparando esta nova técnica com outros métodos de mapeamentos de QTLs, adotando outros delineamentos experimentais ou outras populações, bem como em *genome wide association studies* para tentar reduzir a dimensionalidade das matrizes de genótipos.

REFERÊNCIAS

BALESTRE, M. et al. Bayesian mapping of multiple traits in maize: the importance of pleiotropic effects in studying the inheritance of quantitative traits. **Theoretical and Applied Genetics**, Berlin, v. 125, n. 3, p. 4479-493, 2012.

CAMP, N. J.; COX, A. **Quantitative trait loci: methods and protocols**. Totowa: Humana, 2002. 359 p. (Methods in Molecular Biology, 195).

COCKERHAM, C. C. An extension of the concept of partitioning hereditary variance for analysis of covariance among relatives when epistasis is present. **Genetics**, Baltimore, v. 39, n. 6, p. 859-882, 1954.

DOERGE, R. W.; ZENG, Z. B.; WEIR, B. S. Statistical issues in the search for genes affecting quantitative traits in experimental populations. **Statistical Science**, Hayward, v. 12, n. 3, p. 195-219, 1997.

EDWARDS, M. D.; STUBER, C. W.; WENDEL, J. F. Molecular-marker-facilitated investigations of quantitative-trait loci in maize: I. numbers, genomic distribution and types of gene action. **Genetics**, Baltimore, v. 116, p. 113-125, 1987.

HALDANE, J. B. S. The combination of linkage values and the calculation of distance between the loci of linked factors. **Journal of Genetics**, Bangalore, v. 8, p. 299-309, 1919.

HASTINGS, W. K. Monte Carlo sampling methods using Markov chains and their applications. **Biometrika**, London, v. 57, n. 1, p. 97-109, 1970.

JANSEN, R. C. Interval mapping of multiple quantitative trait loci. **Genetics**, Baltimore, v. 135, n. 1, p. 205-211, 1993.

JOEHANES, R.; NELSON, J. C. QGene 4.0, an extensible Java QTL-analysis platform. **Bioinformatics**, Oxford, v. 24, n. 23, p. 2788-2789, 2008.

KAO, C. H.; ZENG, Z. B.; TEASDALE, R. D. Multiple interval mapping for quantitative trait loci. **Genetics**, Baltimore, v. 152, n. 3, p. 1203-1216, 1999.

KOSAMBI, D. D. The estimation of map distances from recombination values.

Annals of Eugenics, London, v. 12, n. 1, p. 172-175, 1944.

LANDER, E. S.; BOTSTEIN, D. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. **Genetics**, Baltimore, v. 121, n. 1, p. 185-199, 1989.

LARA, L. A. C. et al. Identification of QTLs for resistance to *Sclerotinia sclerotiorum* in Carioca common bean by the moving away method. **ISRN Molecular Biology**, Cairo, 2014.

MARTINEZ, O.; CURNOW, R. N. Estimating the locations and the size of the effects of quantitative trait loci using flanking markers. **Theoretical and Applied Genetics**, Berlin, v. 85, n. 4, p. 480-488, 1992.

METROPOLIS, N. et al. Equations of state calculations by fast computing machines. **Journal of Chemical Physics**, New York, v. 21, n. 6, p. 1087-1092, 1953.

MORAES, M. C. **Mapas de ligação e mapeamento de QTL ("Quantitative Trait Loci") em maracujá-amarelo (*Passiflora edulis* Sims f. *flavicarpa* Deg.)**. 2005. 141 p. Tese (Doutorado em Genética e Melhoramento de Plantas) - Escola Superior de Agricultura "Luiz de Queiroz", Piracicaba, 2005.

R CORE TEAM. **R: a language and environment for statistical computing**. Vienna: R Foundation for Statistical Computing, 2014. Disponível em: <<http://www.R-project.org/>>. Acesso em: 10 maio 2014.

TOLEDO, E. R. et al. Mapeamento de QTLs: uma abordagem bayesiana. **Revista Brasileira de Biometria**, São Paulo, v. 26, n. 2, p. 107-114, 2008.

WANG, H. et al. Bayesian shrinkage estimation of quantitative trait loci parameters. **Genetics**, Baltimore, v. 170, n. 1, p. 465-480, 2005.

WU, R.; MA, C. X.; CASELLA, G. **Statistical genetics of quantitative traits: linkage, maps and QTL**. Berlin: Springer-Verlag, 2007. 361 p.

XU, S. Estimating polygenic effects using markers of the entire genome. **Genetics**, Baltimore, v. 163, n. 2, p. 789-801, 2003.

YANG, R.; XU, S. Bayesian shrinkage analysis of quantitative trait loci for dynamic traits. **Genetics**, Baltimore, v. 176, n. 2, p. 1169-1185, 2007.

ZENG, Z. B. Precision mapping of quantitative trait loci. **Genetics**, Baltimore, v. 136, n. 4, p. 1457-1468, 1994.

ZENG, Z. B. Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci. **Proceedings of the National Academic of Science of the United State of America**, Washington, v. 90, n. 23, p. 10972-10976, 1993.

APÊNDICES

APÊNDICE A. A seguir são apresentados os gráficos de poder em determinar efeitos dominantes, para o Método I (em vermelho) e para o Método II (em azul).

1 Para perda de 20% das marcas.

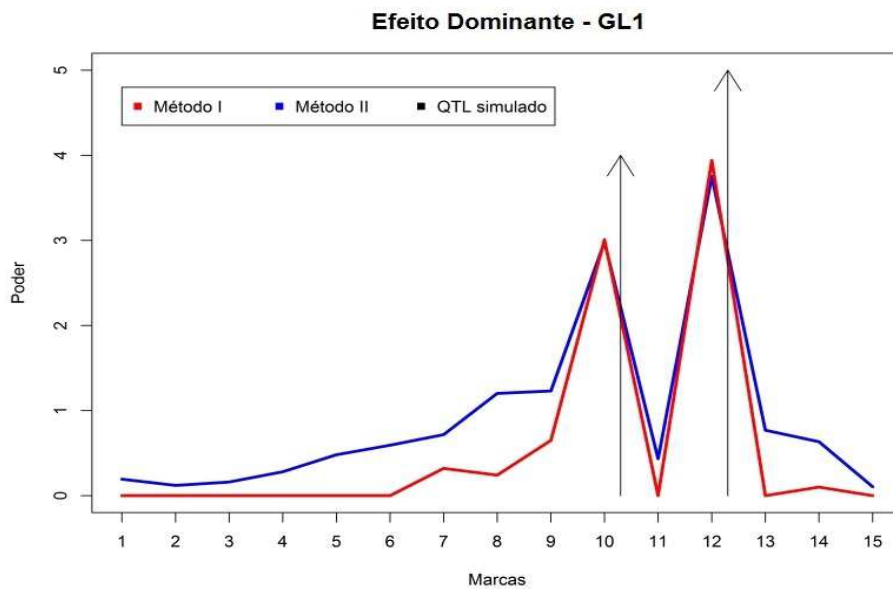


Figura 34 Poder dos Métodos I e II para determinar efeito dominante dos QTLs, para o primeiro grupo de ligação simulado, com 20% de perda de marcadores.

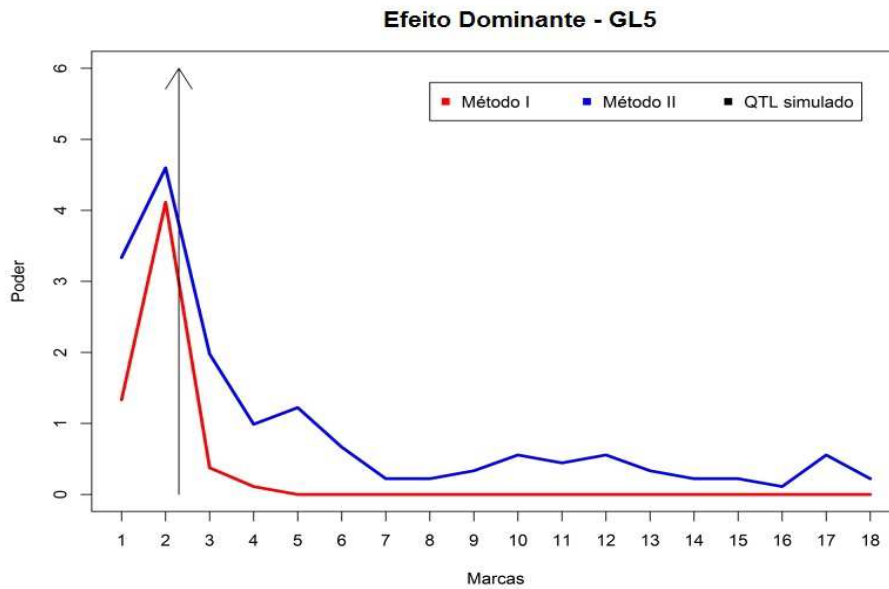


Figura 35 Poder dos Métodos I e II para determinar efeito dominante dos QTLs, para o quinto grupo de ligação simulado, com 20% de perda de marcadores.

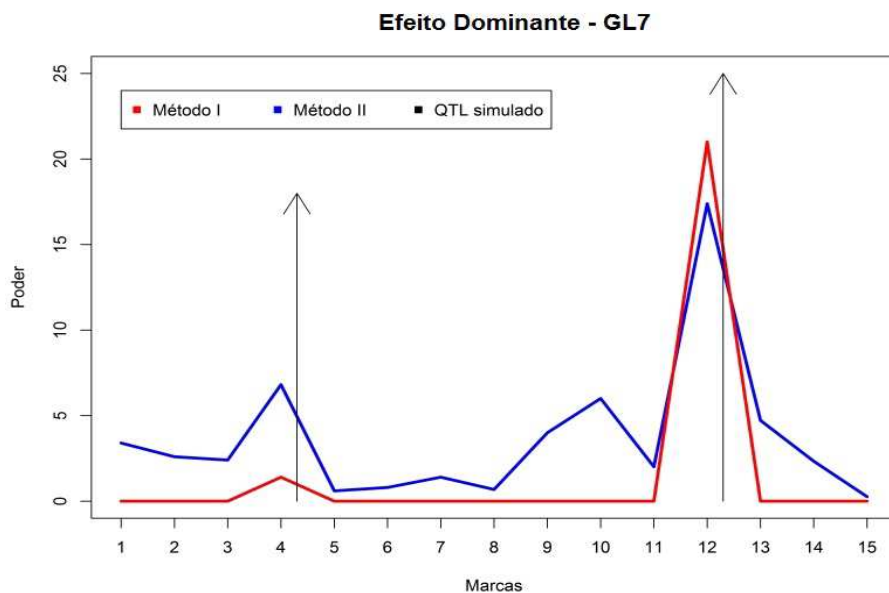


Figura 36 Poder dos Métodos I e II para determinar efeito dominante dos QTLs, para o sétimo grupo de ligação simulado, com 20% de perda de marcadores.

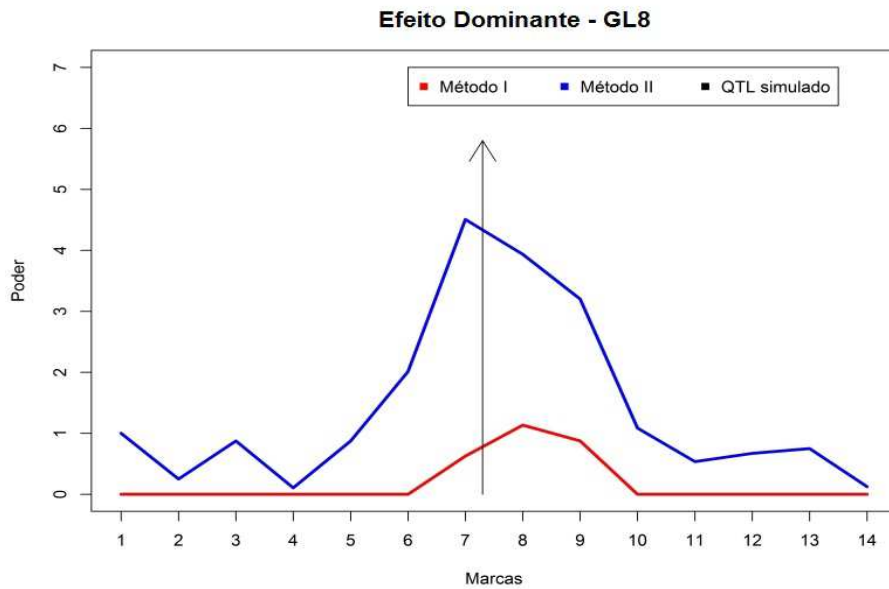


Figura 37 Poder dos Métodos I e II para determinar efeito dominante dos QTLs, para o oitavo grupo de ligação simulado, com 20% de perda de marcadores.

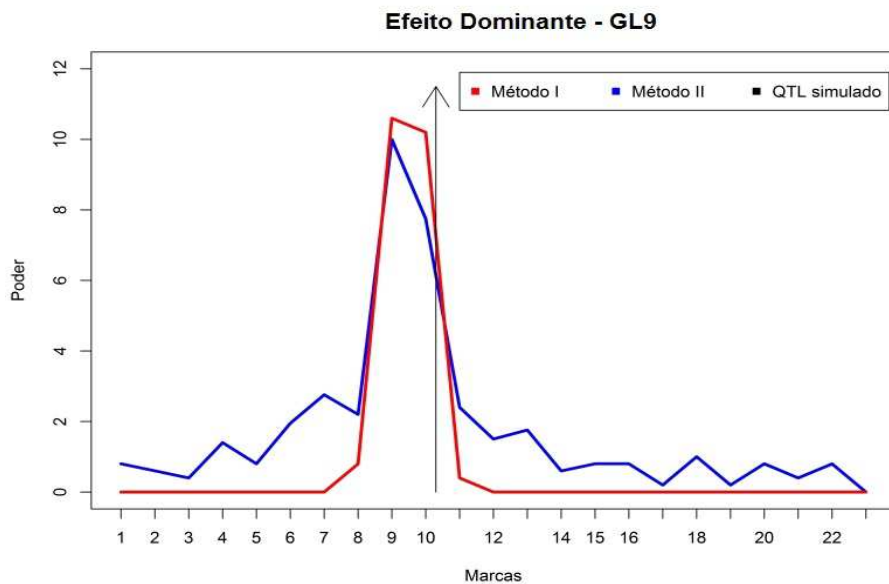


Figura 38 Poder dos Métodos I e II para determinar efeito dominante dos QTLs, para o nono grupo de ligação simulado, com 20% de perda de marcadores.

2 Para perda de 80% das marcas.

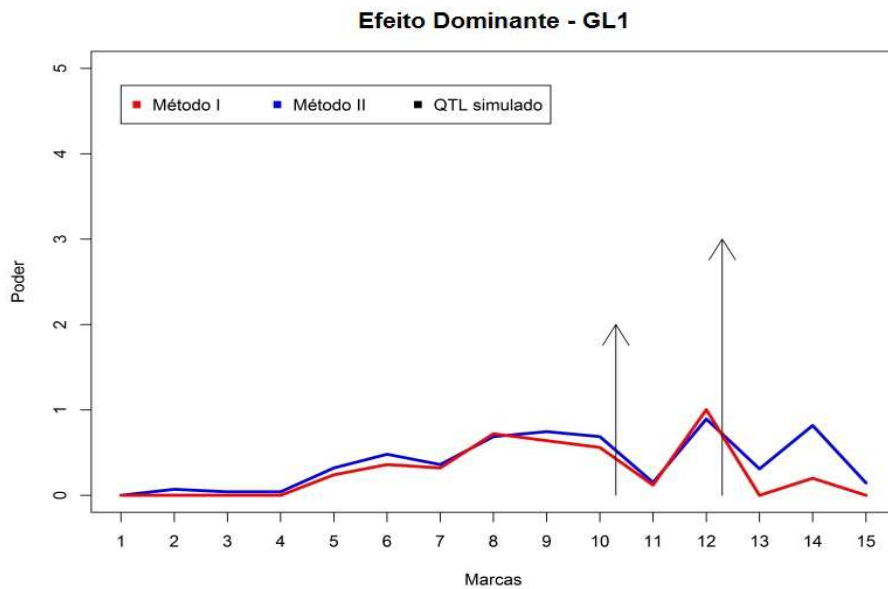


Figura 39 Poder dos Métodos I e II para determinar efeito dominante dos QTLs, para o primeiro grupo de ligação simulado, com 80% de perda de marcadores.

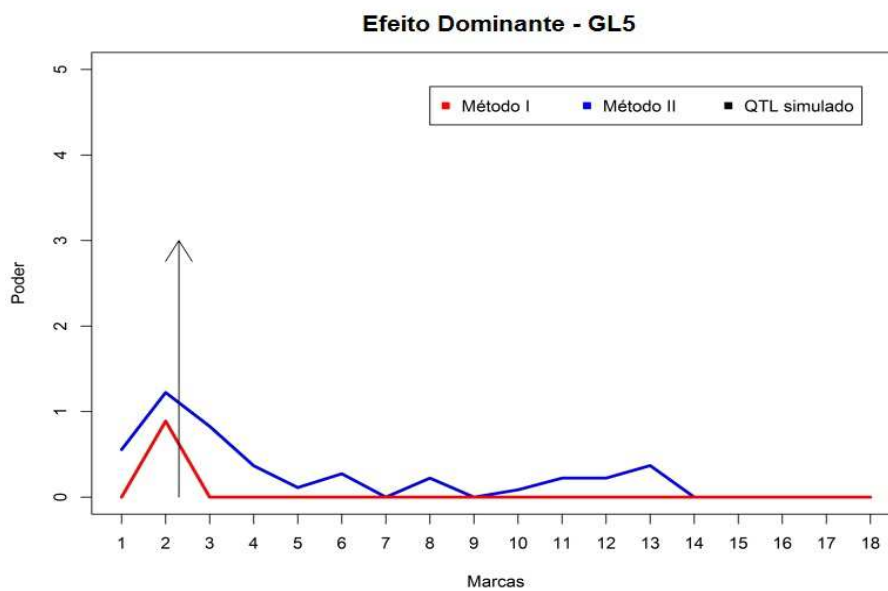


Figura 40 Poder dos Métodos I e II para determinar efeito dominante dos QTLs, para o quinto grupo de ligação simulado, com 80% de perda de marcadores.

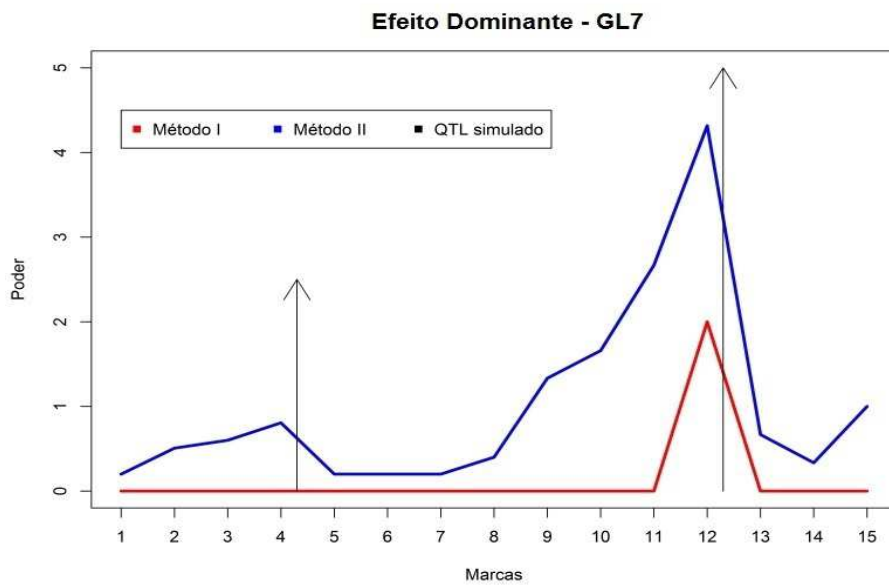


Figura 41 Poder dos Métodos I e II para determinar efeito dominante dos QTLs, para o sétimo grupo de ligação simulado, com 80% de perda de marcadores.

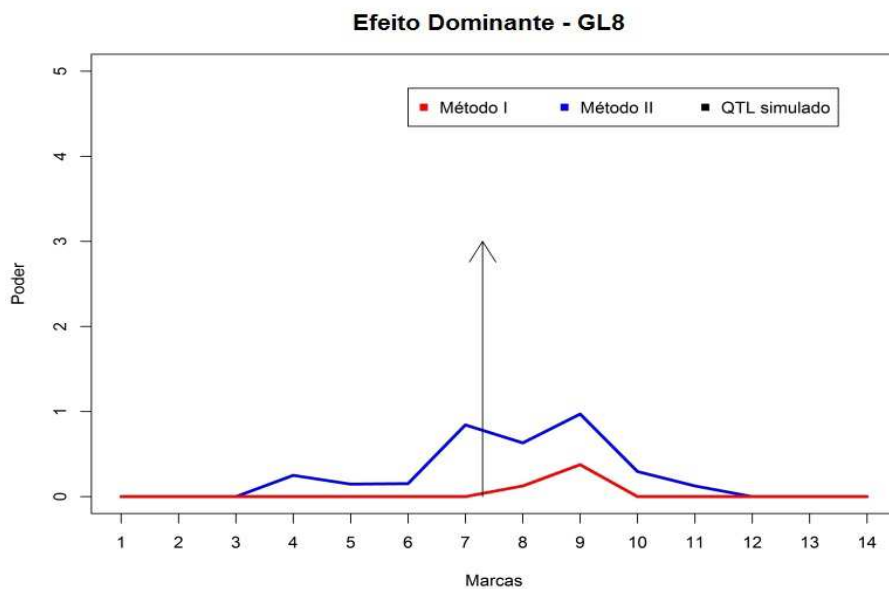


Figura 42 Poder dos Métodos I e II para determinar efeito dominante dos QTLs, para o oitavo grupo de ligação simulado, com 80% de perda de marcadores.

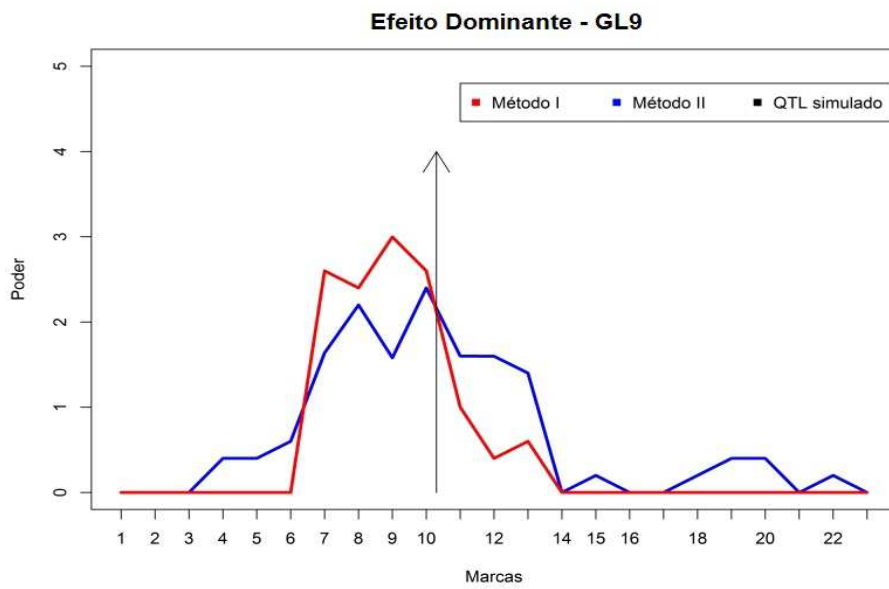


Figura 43 Poder dos Métodos I e II para determinar efeito dominante dos QTLs, para o nono grupo de ligação simulado, com 80% de perda de marcadores.

APÊNDICE B. A seguir são apresentados os gráficos com as posições dos QTLs estimados por ambos os métodos, as frequências relativas com os respectivos intervalos de confiança exatos para proporção com 5% de significância e as diferenças médias de efeitos dominantes entre o QTL estimado e o QTL simulado.

- 1 Para perda de 20% das marcas.

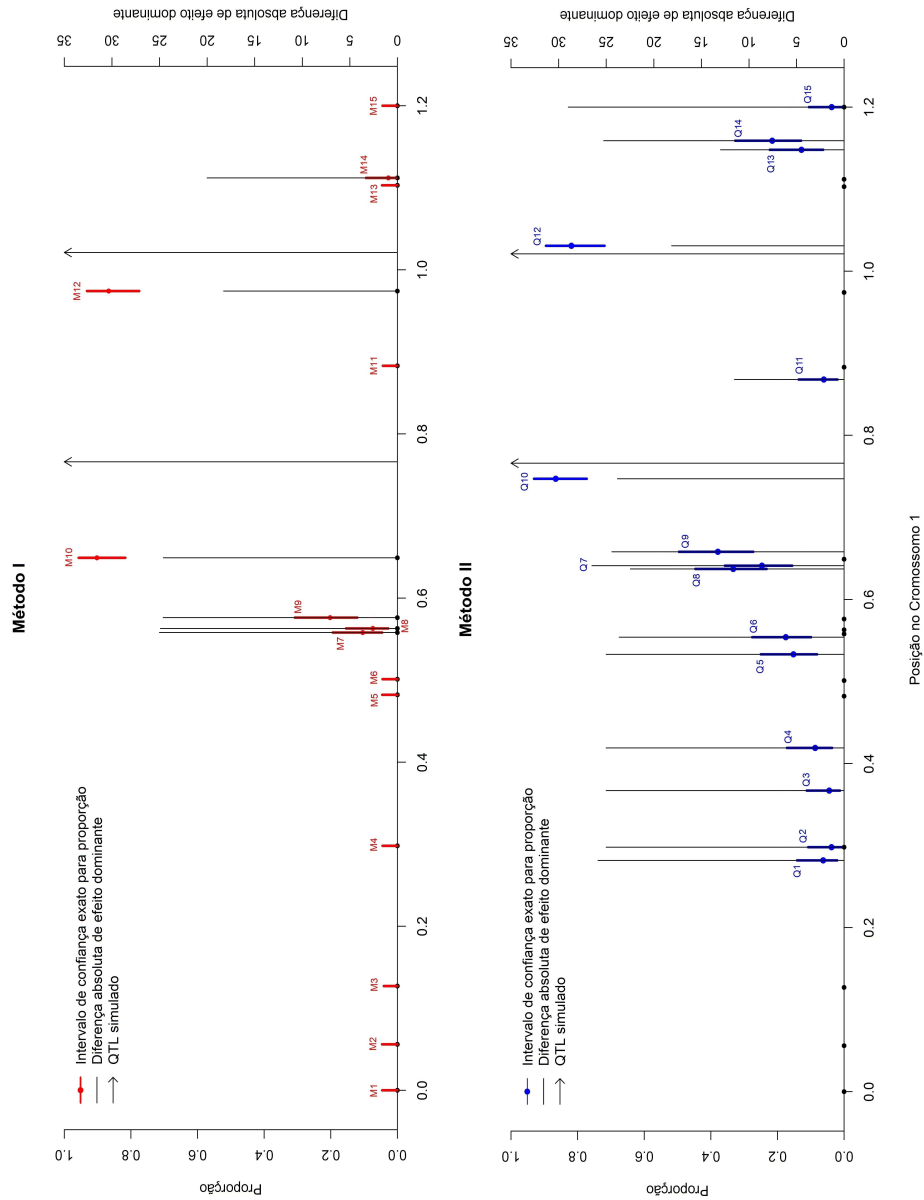


Figura 44 QTLs estimados pelos Métodos I e II com as respectivas posições (em unidade Morgan) no cromossomo um, frequências relativas com intervalos de confiança exato para proporção com 5% de significância e diferenças absolutas dos efeitos dominantes entre os QTLs estimados e os QTLs simulados, para perda de 20% das marcas.

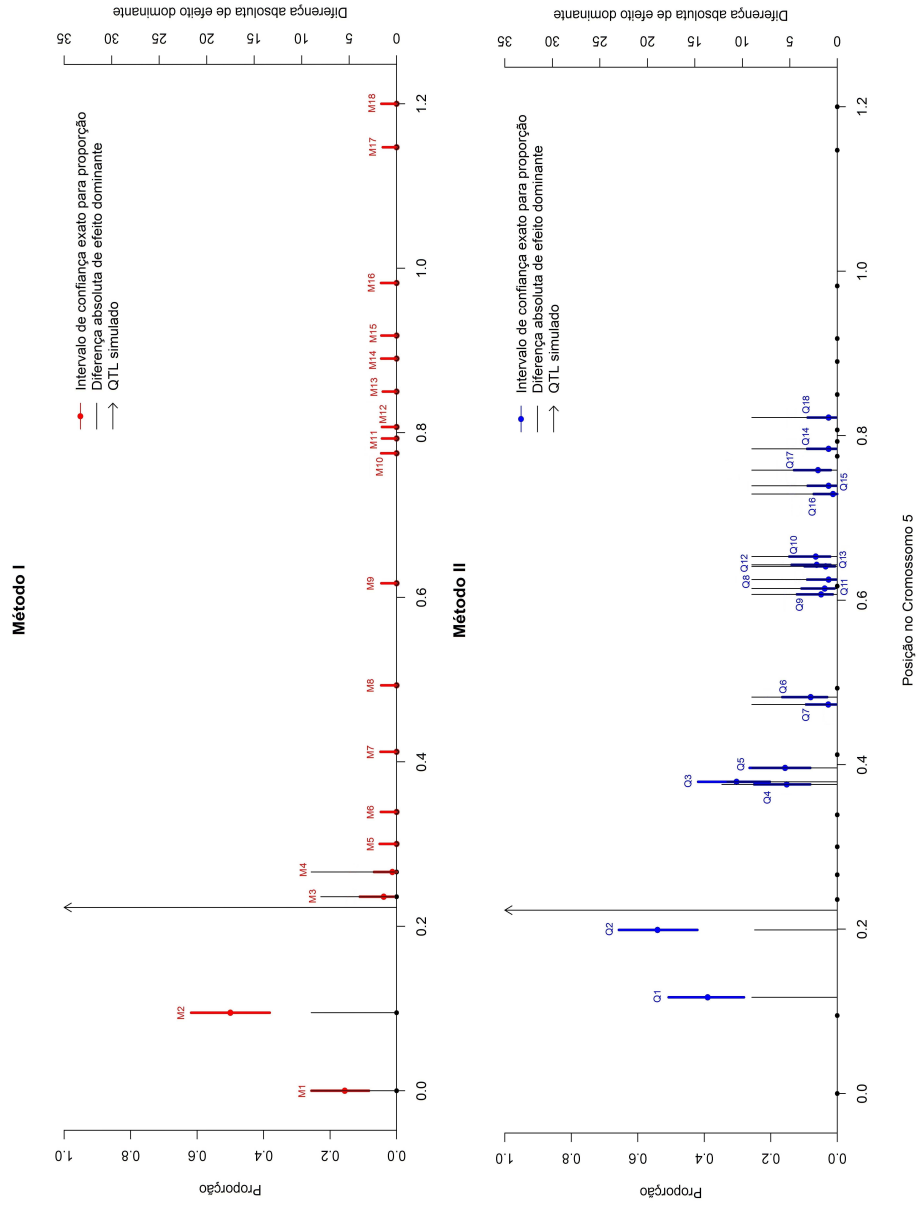


Figura 45 QTLs estimados pelos Métodos I e II com as respectivas posições (em unidade Morgan) no cromossomo cinco, frequências relativas com intervalos de confiança exato para proporção com 5% de significância e diferenças absolutas dos efeitos dominantes entre os QTLs estimados e os QTLs simulados, para perda de 20% das marcas.

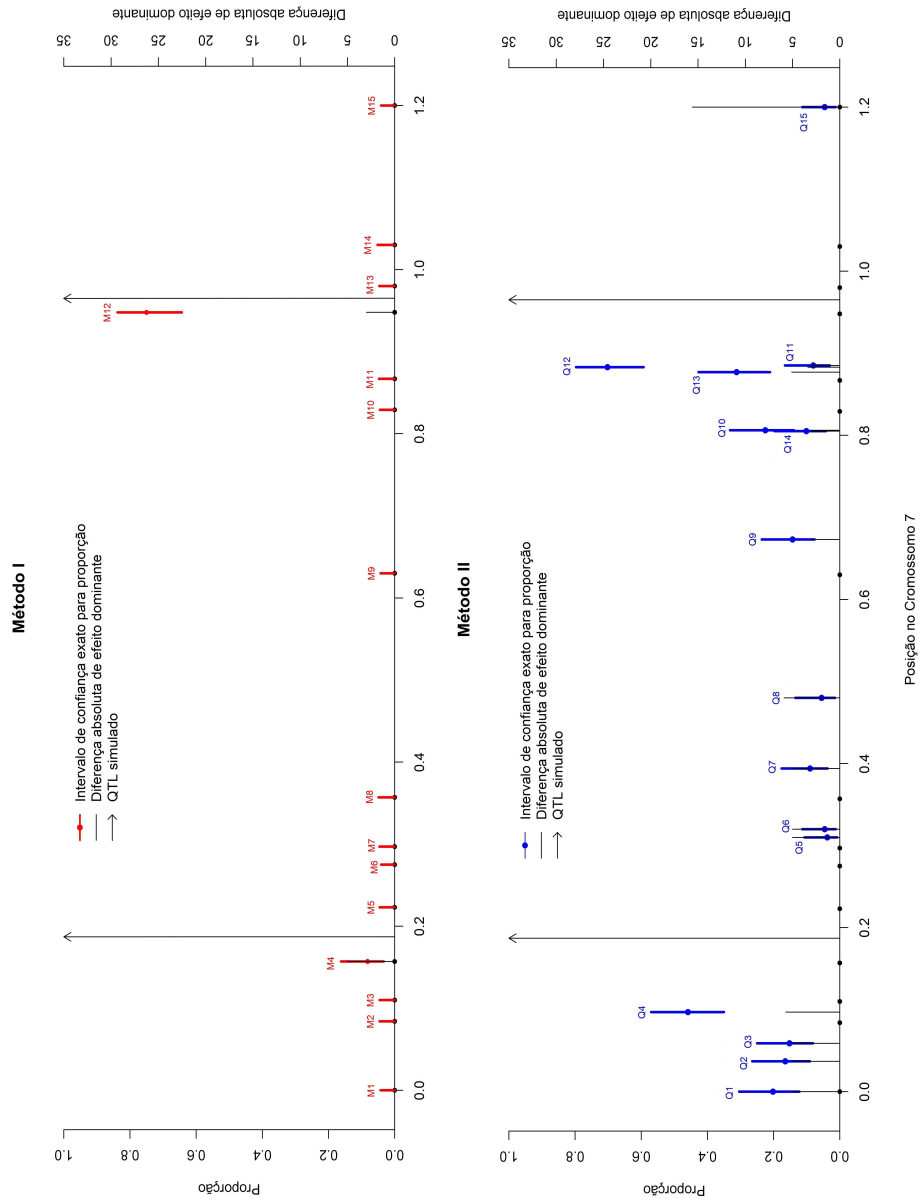


Figura 46 QTLs estimados pelos Métodos I e II com as respectivas posições (em unidade Morgan) no cromossomo sete, frequências relativas com intervalos de confiança exato para proporção com 5% de significância e diferenças absolutas dos efeitos dominantes entre os QTLs estimados e os QTLs simulados, para perda de 20% das marcas.

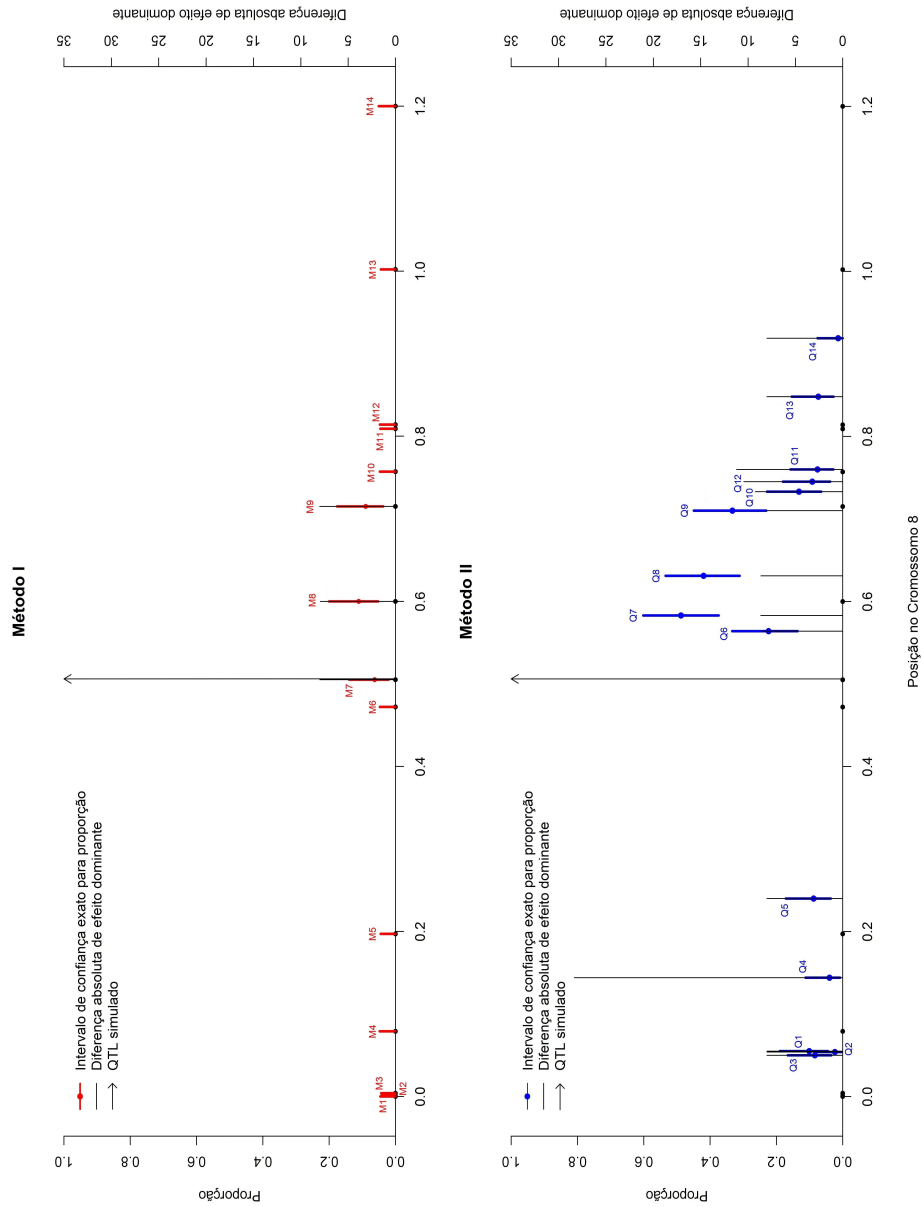


Figura 47 QTLs estimados pelos Métodos I e II com as respectivas posições (em unidade Morgan) no cromossomo oito, frequências relativas com intervalos de confiança exato para proporção com 5% de significância e diferenças absolutas dos efeitos dominantes entre os QTLs estimados e os QTLs simulados, para perda de 20% das marcas.

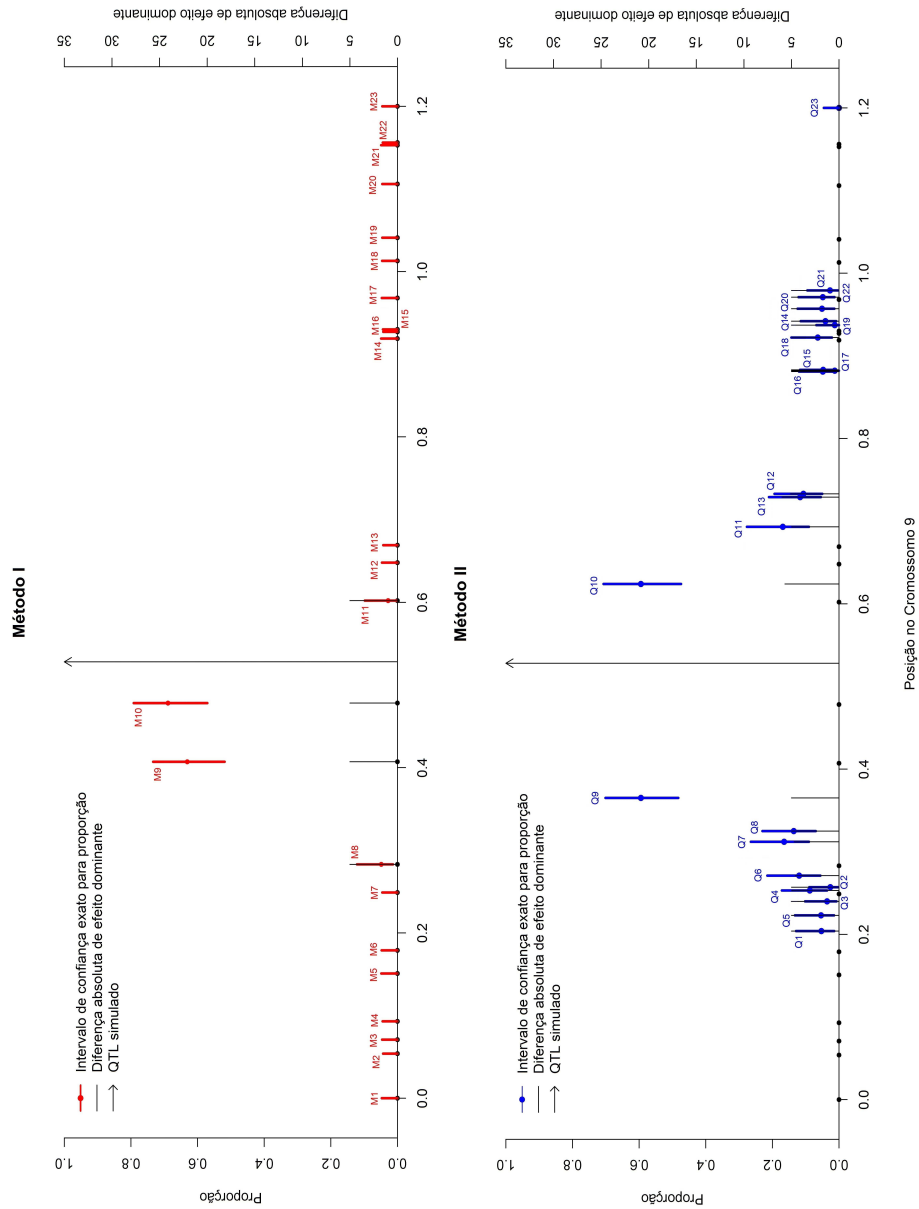


Figura 48 QTLs estimados pelos Métodos I e II com as respectivas posições (em unidade Morgan) no cromossomo nove, frequências relativas com intervalos de confiança exato para proporção com 5% de significância e diferenças absolutas dos efeitos dominantes entre os QTLs estimados e os QTLs simulados, para perda de 20% das marcas.

2 Para perda de 80% das marcas.

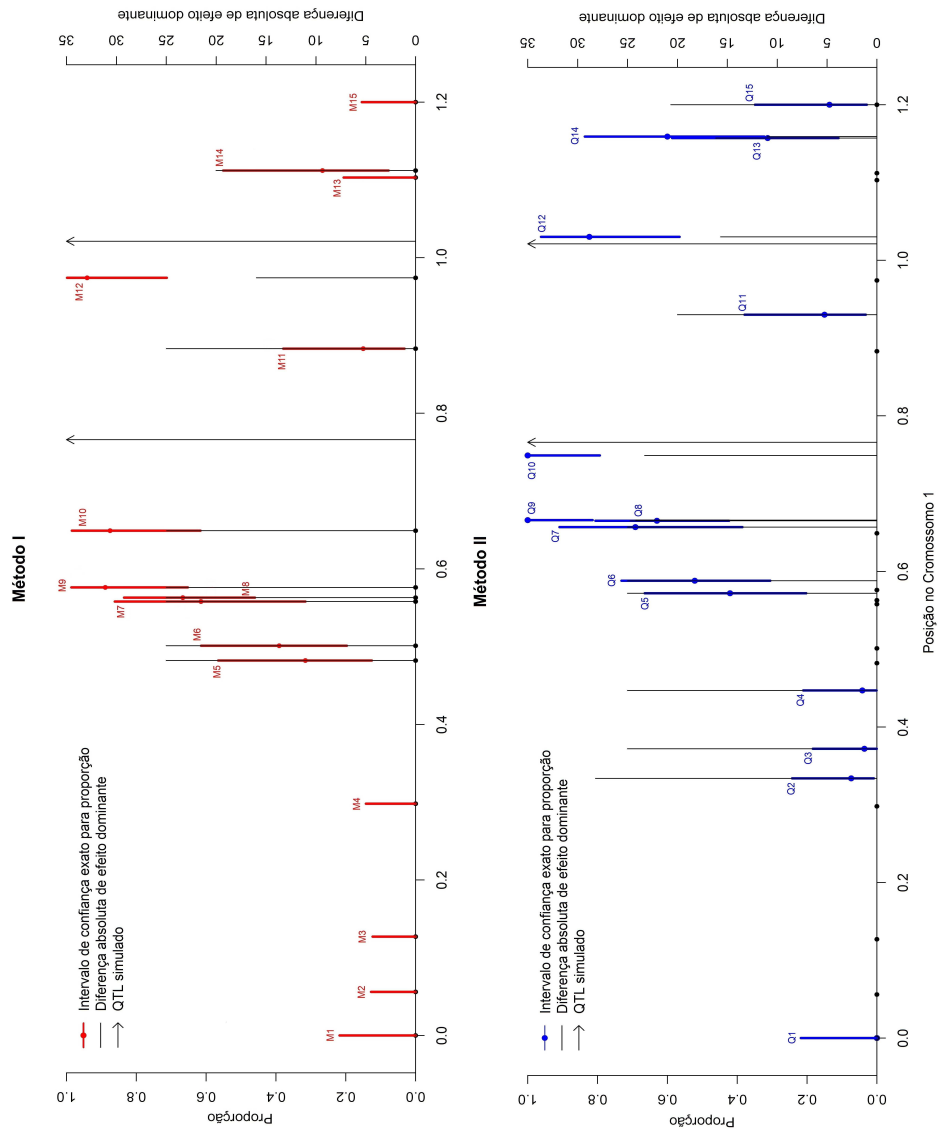


Figura 49 QTLs estimados pelos Métodos I e II com as respectivas posições (em unidade Morgan) no cromossomo um, frequências relativas com intervalos de confiança exato para proporção com 5% de significância e diferenças absolutas dos efeitos dominantes entre os QTLs estimados e os QTLs simulados, para perda de 80% das marcas.

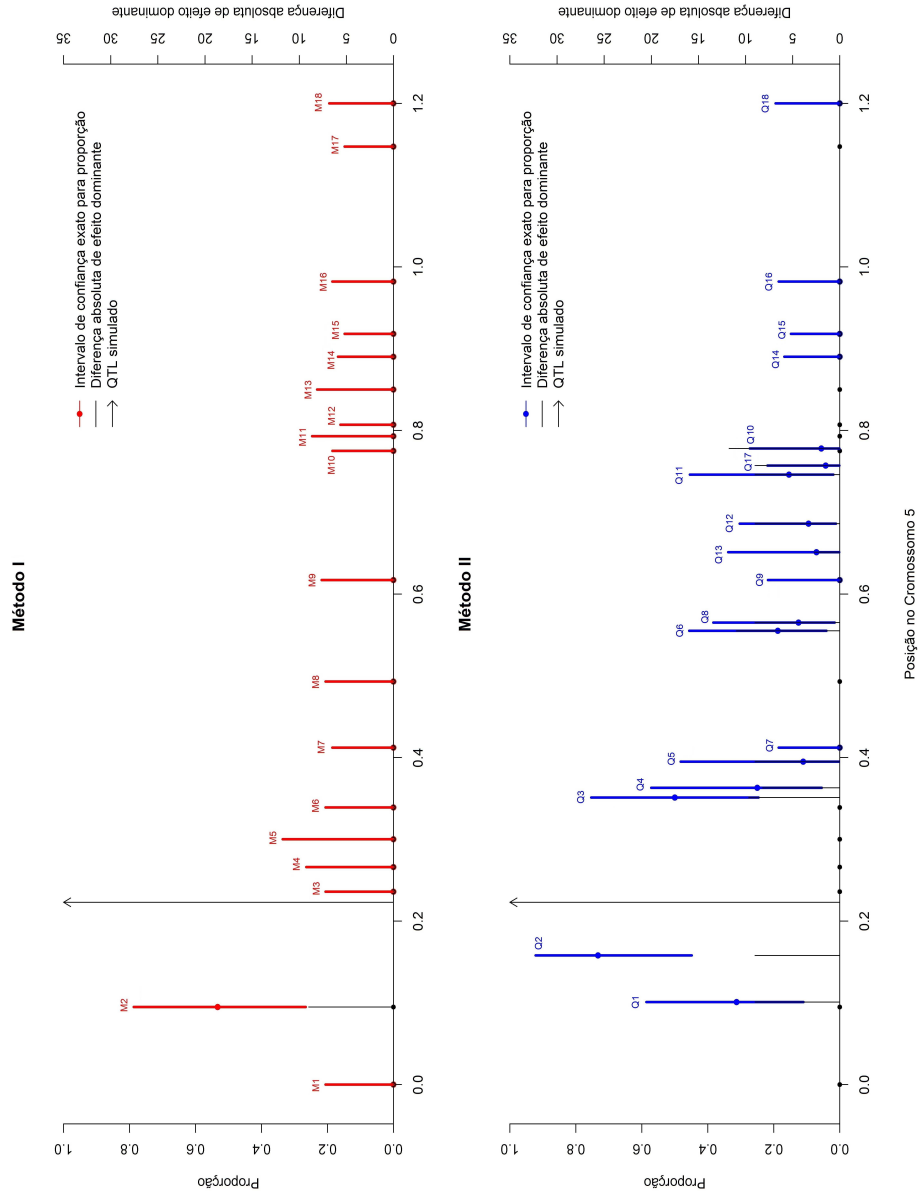


Figura 50 QTLs estimados pelos Métodos I e II com as respectivas posições (em unidade Morgan) no cromossomo cinco, frequências relativas com intervalos de confiança exato para proporção com 5% de significância e diferenças absolutas dos efeitos dominantes entre os QTLs estimados e os QTLs simulados, para perda de 80% das marcas.

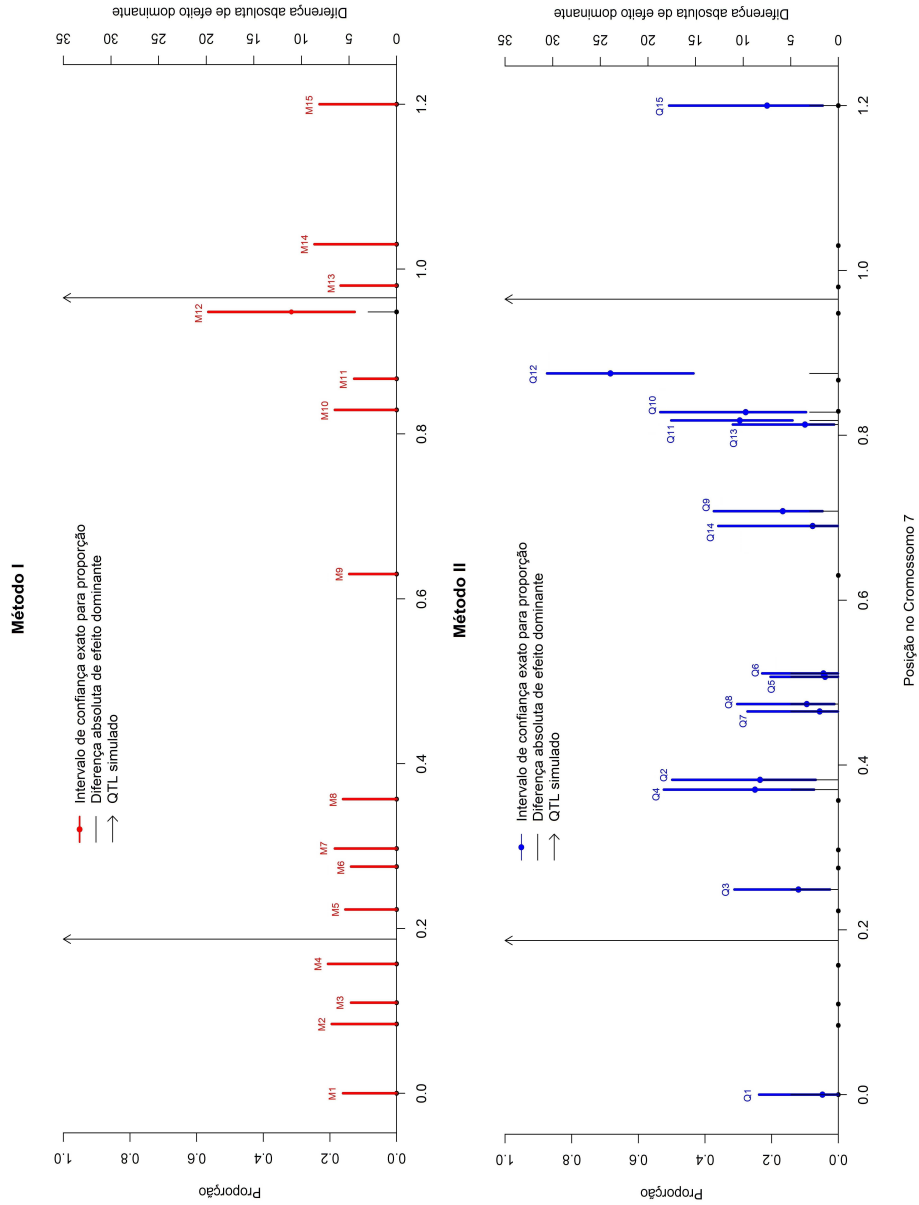


Figura 51 QTLs estimados pelos Métodos I e II com as respectivas posições (em unidade Morgan) no cromossomo sete, frequências relativas com intervalos de confiança exato para proporção com 5% de significância e diferenças absolutas dos efeitos dominantes entre os QTLs estimados e os QTLs simulados, para perda de 80% das marcas.

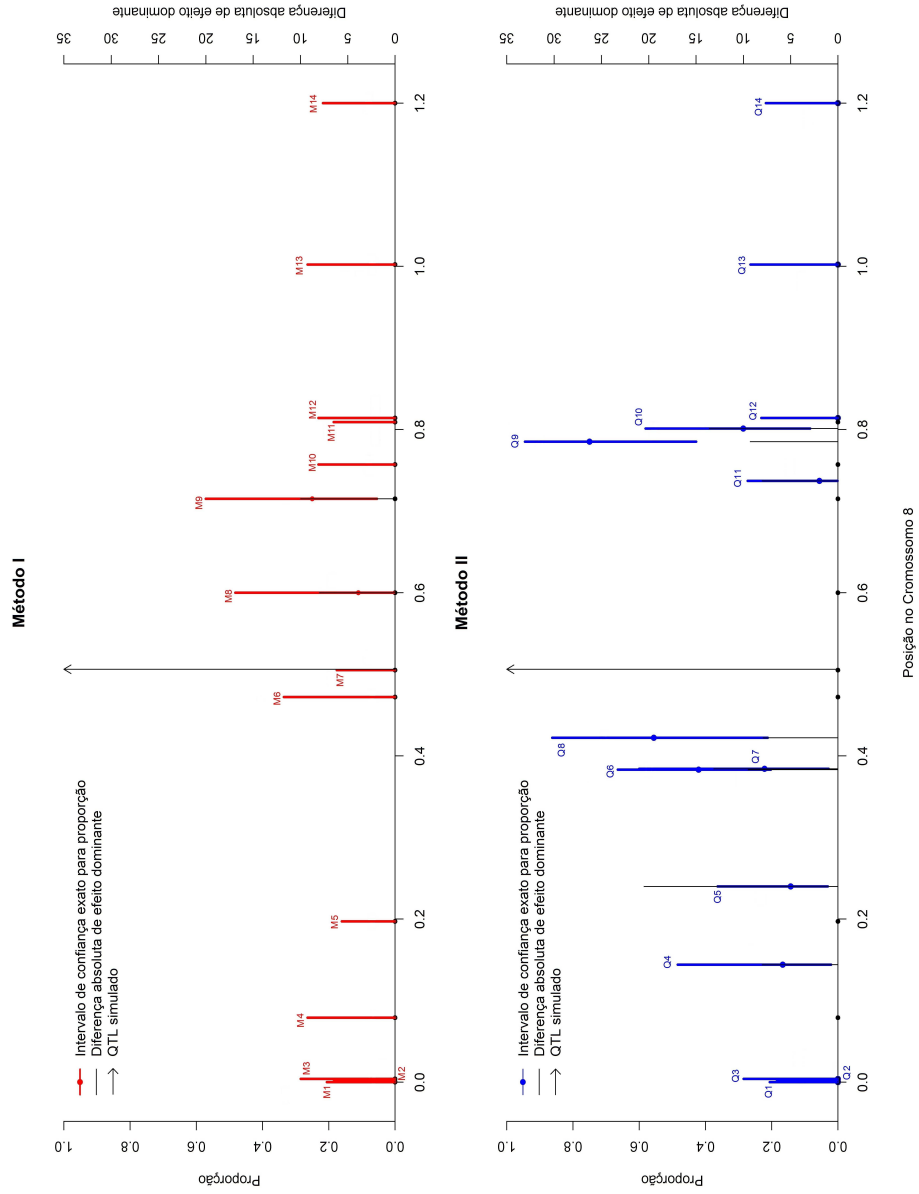


Figura 52 QTLs estimados pelos Métodos I e II com as respectivas posições (em unidade Morgan) no cromossomo oito, frequências relativas com intervalos de confiança exato para proporção com 5% de significância e diferenças absolutas dos efeitos dominantes entre os QTLs estimados e os QTLs simulados, para perda de 80% das marcas.

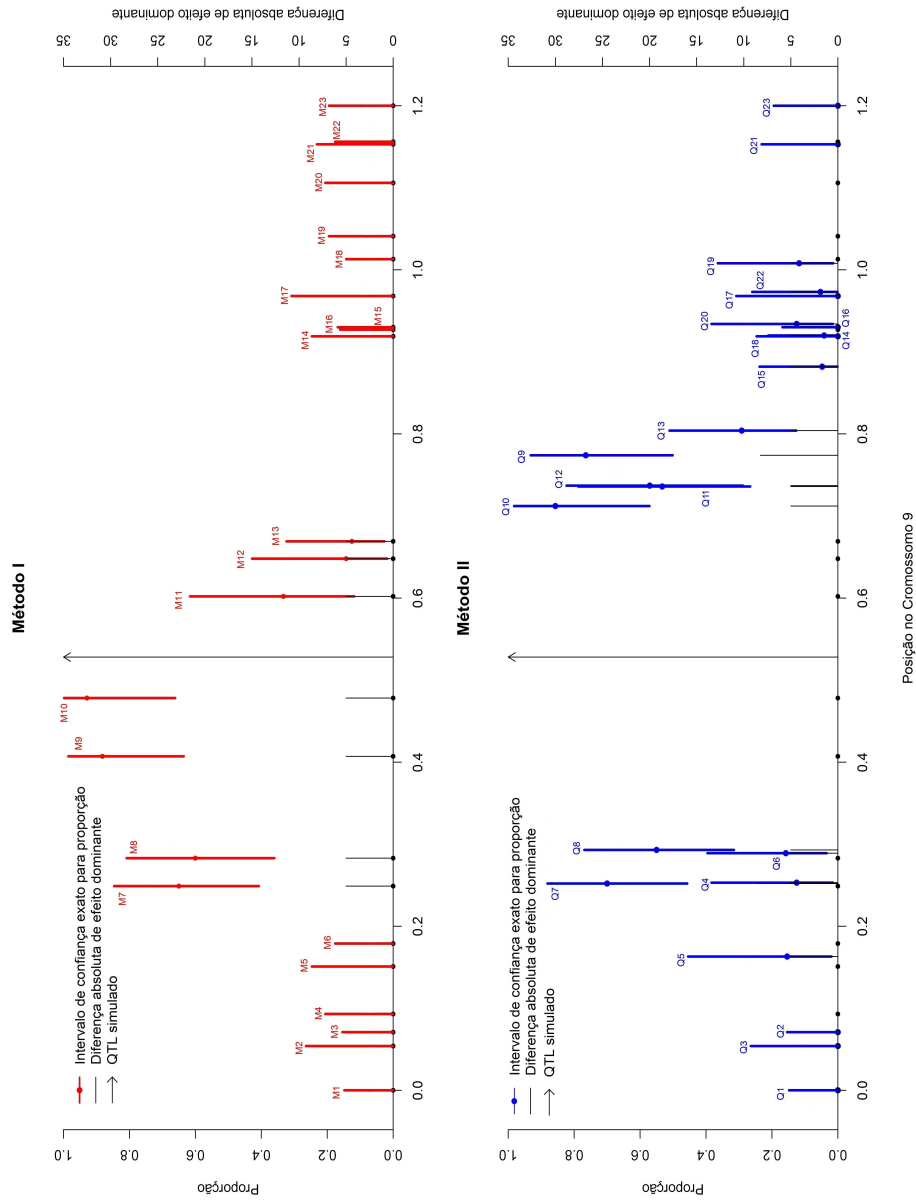


Figura 53 QTLs estimados pelos Métodos I e II com as respectivas posições (em unidade Morgan) no cromossomo nove, frequências relativas com intervalos de confiança exato para proporção com 5% de significância e diferenças absolutas dos efeitos dominantes entre os QTLs estimados e os QTLs simulados, para perda de 80% das marcas.

APÊNDICE C. A seguir são apresentadas as tabelas com os resumos dos resultados obtidos em cada método, para ambos os níveis de perdas de marcadores.

Tabela 2 Frequências de detecção, total de seleções e médias das distâncias para os dois métodos, por grupo de ligação (GL), com perda de 20% das marcas.

| GL | Marcador | Total de Seleções | Método I | | Método II | |
|----|----------|-------------------|---------------------|-----------------|---------------------|-----------------|
| | | | Frequência Detecção | Distância Média | Frequência Detecção | Distância Média |
| 1 | 1 | 79 | 0 | - | 5 | 0,484 |
| 1 | 2 | 78 | 0 | - | 3 | 0,468 |
| 1 | 3 | 88 | 0 | - | 4 | 0,399 |
| 1 | 4 | 80 | 0 | - | 7 | 0,347 |
| 1 | 5 | 79 | 0 | - | 12 | 0,233 |
| 1 | 6 | 80 | 0 | - | 14 | 0,212 |
| 1 | 7 | 77 | 8 | 0,208 | 19 | 0,125 |
| 1 | 8 | 81 | 6 | 0,203 | 27 | 0,129 |
| 1 | 9 | 79 | 16 | 0,190 | 30 | 0,108 |
| 1 | 10 | 82 | 74 | 0,117 | 71 | 0,019 |
| 1 | 11 | 82 | 0 | - | 5 | 0,153 |
| 1 | 12 | 83 | 72 | 0,047 | 68 | 0,010 |
| 1 | 13 | 78 | 0 | - | 10 | 0,127 |
| 1 | 14 | 74 | 2 | 0,091 | 16 | 0,138 |
| 1 | 15 | 80 | 0 | - | 3 | 0,186 |
| 5 | 1 | 77 | 12 | 0,223 | 30 | 0,106 |
| 5 | 2 | 74 | 37 | 0,128 | 40 | 0,024 |
| 5 | 3 | 76 | 3 | 0,013 | 23 | 0,156 |
| 5 | 4 | 79 | 1 | 0,043 | 12 | 0,153 |
| 5 | 5 | 70 | 0 | - | 11 | 0,173 |
| 5 | 6 | 75 | 0 | - | 6 | 0,259 |
| 5 | 7 | 74 | 0 | - | 2 | 0,250 |
| 5 | 8 | 77 | 0 | - | 2 | 0,402 |
| 5 | 9 | 78 | 0 | - | 3 | 0,391 |
| 5 | 10 | 77 | 0 | - | 5 | 0,430 |

| GL | Marcador | Total de Seleções | Método I | | Método II | |
|----|----------|-------------------|---------------------|-----------------|---------------------|-----------------|
| | | | Frequência Detecção | Distância Média | Frequência Detecção | Distância Média |
| 5 | 11 | 81 | 0 | - | 4 | 0,384 |
| 5 | 12 | 81 | 0 | - | 5 | 0,420 |
| 5 | 13 | 86 | 0 | - | 3 | 0,418 |
| 5 | 14 | 77 | 0 | - | 2 | 0,561 |
| 5 | 15 | 78 | 0 | - | 2 | 0,516 |
| 5 | 16 | 76 | 0 | - | 1 | 0,506 |
| 5 | 17 | 86 | 0 | - | 5 | 0,535 |
| 5 | 18 | 78 | 0 | - | 2 | 0,599 |
| 7 | 1 | 84 | 0 | - | 17 | 0,213 |
| 7 | 2 | 79 | 0 | - | 13 | 0,150 |
| 7 | 3 | 79 | 0 | - | 12 | 0,128 |
| 7 | 4 | 85 | 7 | 0,030 | 39 | 0,090 |
| 7 | 5 | 79 | 0 | - | 3 | 0,123 |
| 7 | 6 | 87 | 0 | - | 4 | 0,133 |
| 7 | 7 | 78 | 0 | - | 7 | 0,207 |
| 7 | 8 | 73 | 0 | - | 4 | 0,293 |
| 7 | 9 | 84 | 0 | - | 12 | 0,292 |
| 7 | 10 | 80 | 0 | - | 18 | 0,159 |
| 7 | 11 | 75 | 0 | - | 6 | 0,080 |
| 7 | 12 | 84 | 63 | 0,017 | 59 | 0,082 |
| 7 | 13 | 77 | 0 | - | 24 | 0,088 |
| 7 | 14 | 69 | 0 | - | 7 | 0,160 |
| 7 | 15 | 87 | 0 | - | 4 | 0,244 |
| 8 | 1 | 79 | 0 | - | 8 | 0,451 |
| 8 | 2 | 85 | 0 | - | 2 | 0,452 |
| 8 | 3 | 83 | 0 | - | 7 | 0,456 |
| 8 | 4 | 75 | 0 | - | 3 | 0,362 |
| 8 | 5 | 80 | 0 | - | 7 | 0,266 |
| 8 | 6 | 76 | 0 | - | 17 | 0,058 |
| 8 | 7 | 80 | 5 | 0,001 | 39 | 0,077 |
| 8 | 8 | 81 | 9 | 0,094 | 34 | 0,125 |
| 8 | 9 | 78 | 7 | 0,209 | 26 | 0,204 |
| 8 | 10 | 76 | 0 | - | 10 | 0,227 |
| 8 | 11 | 79 | 0 | - | 6 | 0,254 |
| 8 | 12 | 76 | 0 | - | 7 | 0,239 |

| GL | Marcador | Total de Seleções | Método I | | Método II | |
|----|----------|-------------------|---------------------|-----------------|---------------------|-----------------|
| | | | Frequência Detecção | Distância Média | Frequência Detecção | Distância Média |
| 8 | 13 | 81 | 0 | - | 6 | 0,342 |
| 8 | 14 | 71 | 0 | - | 1 | 0,413 |
| 9 | 1 | 76 | 0 | - | 4 | 0,324 |
| 9 | 2 | 83 | 0 | - | 3 | 0,288 |
| 9 | 3 | 78 | 0 | - | 2 | 0,271 |
| 9 | 4 | 80 | 0 | - | 7 | 0,275 |
| 9 | 5 | 74 | 0 | - | 4 | 0,305 |
| 9 | 6 | 75 | 0 | - | 9 | 0,257 |
| 9 | 7 | 79 | 0 | - | 13 | 0,216 |
| 9 | 8 | 81 | 4 | 0,245 | 11 | 0,203 |
| 9 | 9 | 84 | 53 | 0,121 | 50 | 0,163 |
| 9 | 10 | 74 | 51 | 0,050 | 44 | 0,096 |
| 9 | 11 | 71 | 2 | 0,074 | 12 | 0,165 |
| 9 | 12 | 77 | 0 | - | 9 | 0,201 |
| 9 | 13 | 84 | 0 | - | 9 | 0,205 |
| 9 | 14 | 73 | 0 | - | 3 | 0,414 |
| 9 | 15 | 84 | 0 | - | 4 | 0,355 |
| 9 | 16 | 82 | 0 | - | 4 | 0,353 |
| 9 | 17 | 76 | 0 | - | 1 | 0,354 |
| 9 | 18 | 78 | 0 | - | 5 | 0,394 |
| 9 | 19 | 79 | 0 | - | 1 | 0,409 |
| 9 | 20 | 79 | 0 | - | 4 | 0,429 |
| 9 | 21 | 73 | 0 | - | 2 | 0,451 |
| 9 | 22 | 81 | 0 | - | 4 | 0,443 |
| 9 | 23 | 79 | 0 | - | 0 | - |

Tabela 3 Frequências de detecção, total de seleções e médias das distâncias para os dois métodos, por grupo de ligação (GL), com perda de 80% das marcas.

| GL | Marcador | Total de Seleções | Método I | | Método II | |
|----|----------|-------------------|---------------------|-----------------|---------------------|-----------------|
| | | | Frequência Detecção | Distância Média | Frequência Detecção | Distância Média |
| 1 | 1 | 15 | 0 | - | 0 | - |
| 1 | 2 | 27 | 0 | - | 2 | 0,432 |
| 1 | 3 | 28 | 0 | - | 1 | 0,394 |
| 1 | 4 | 24 | 0 | - | 1 | 0,319 |
| 1 | 5 | 19 | 6 | 0,284 | 8 | 0,194 |
| 1 | 6 | 23 | 9 | 0,265 | 12 | 0,178 |
| 1 | 7 | 13 | 8 | 0,208 | 9 | 0,109 |
| 1 | 8 | 27 | 18 | 0,203 | 17 | 0,101 |
| 1 | 9 | 18 | 16 | 0,190 | 18 | 0,100 |
| 1 | 10 | 16 | 14 | 0,117 | 16 | 0,017 |
| 1 | 11 | 20 | 3 | 0,117 | 3 | 0,091 |
| 1 | 12 | 17 | 16 | 0,047 | 14 | 0,009 |
| 1 | 13 | 16 | 0 | - | 5 | 0,136 |
| 1 | 14 | 15 | 4 | 0,091 | 9 | 0,138 |
| 1 | 15 | 22 | 0 | - | 3 | 0,266 |
| 5 | 1 | 16 | 0 | - | 5 | 0,122 |
| 5 | 2 | 15 | 8 | 0,128 | 11 | 0,065 |
| 5 | 3 | 16 | 0 | - | 8 | 0,128 |
| 5 | 4 | 12 | 0 | - | 3 | 0,140 |
| 5 | 5 | 9 | 0 | - | 1 | 0,172 |
| 5 | 6 | 16 | 0 | - | 3 | 0,332 |
| 5 | 7 | 18 | 0 | - | 0 | - |
| 5 | 8 | 16 | 0 | - | 2 | 0,342 |
| 5 | 9 | 15 | 0 | - | 0 | - |
| 5 | 10 | 18 | 0 | - | 1 | 0,555 |
| 5 | 11 | 13 | 0 | - | 2 | 0,523 |
| 5 | 12 | 21 | 0 | - | 2 | 0,463 |
| 5 | 13 | 14 | 0 | - | 1 | 0,428 |
| 5 | 14 | 20 | 0 | - | 0 | - |
| 5 | 15 | 23 | 0 | - | 0 | - |
| 5 | 16 | 18 | 0 | - | 0 | 0 |
| 5 | 17 | 23 | 0 | - | 1 | 0,534 |
| 5 | 18 | 17 | 0 | - | 0 | 0 |

| GL | Marcador | Total de Seleções | Método I | | Método II | |
|----|----------|-------------------|---------------------|-----------------|---------------------|-----------------|
| | | | Frequência Detecção | Distância Média | Frequência Detecção | Distância Média |
| 7 | 1 | 21 | 0 | - | 1 | 0,417 |
| 7 | 2 | 17 | 0 | - | 4 | 0,195 |
| 7 | 3 | 25 | 0 | - | 3 | 0,062 |
| 7 | 4 | 16 | 0 | - | 4 | 0,183 |
| 7 | 5 | 22 | 0 | - | 1 | 0,324 |
| 7 | 6 | 25 | 0 | - | 1 | 0,320 |
| 7 | 7 | 18 | 0 | - | 1 | 0,278 |
| 7 | 8 | 21 | 0 | - | 2 | 0,287 |
| 7 | 9 | 24 | 0 | - | 4 | 0,257 |
| 7 | 10 | 18 | 0 | - | 5 | 0,137 |
| 7 | 11 | 27 | 0 | - | 8 | 0,147 |
| 7 | 12 | 19 | 6 | 0,017 | 13 | 0,090 |
| 7 | 13 | 20 | 0 | - | 2 | 0,152 |
| 7 | 14 | 13 | 0 | - | 1 | 0,275 |
| 7 | 15 | 14 | 0 | - | 3 | 0,286 |
| 8 | 1 | 16 | 0 | - | 0 | - |
| 8 | 2 | 18 | 0 | - | 0 | - |
| 8 | 3 | 11 | 0 | - | 0 | - |
| 8 | 4 | 12 | 0 | - | 2 | 0,362 |
| 8 | 5 | 21 | 0 | - | 3 | 0,266 |
| 8 | 6 | 9 | 0 | - | 2 | 0,122 |
| 8 | 7 | 19 | 0 | - | 8 | 0,123 |
| 8 | 8 | 9 | 1 | 0,094 | 5 | 0,084 |
| 8 | 9 | 12 | 3 | 0,209 | 9 | 0,279 |
| 8 | 10 | 14 | 0 | - | 4 | 0,295 |
| 8 | 11 | 18 | 0 | - | 1 | 0,231 |
| 8 | 12 | 14 | 0 | - | 0 | - |
| 8 | 13 | 12 | 0 | - | 0 | - |
| 8 | 14 | 15 | 0 | - | 0 | - |
| 9 | 1 | 23 | 0 | - | 0 | - |
| 9 | 2 | 12 | 0 | - | 0 | - |
| 9 | 3 | 22 | 0 | - | 0 | - |
| 9 | 4 | 16 | 0 | - | 2 | 0,275 |
| 9 | 5 | 13 | 0 | - | 2 | 0,365 |
| 9 | 6 | 19 | 0 | - | 3 | 0,239 |

| GL | Marcador | Total de Seleções | Método I | | Método II | |
|----|----------|-------------------|---------------------|-----------------|---------------------|-----------------|
| | | | Frequência Detecção | Distância Média | Frequência Detecção | Distância Média |
| 9 | 7 | 20 | 13 | 0,279 | 14 | 0,276 |
| 9 | 8 | 20 | 12 | 0,245 | 11 | 0,235 |
| 9 | 9 | 17 | 15 | 0,121 | 13 | 0,246 |
| 9 | 10 | 14 | 13 | 0,050 | 12 | 0,184 |
| 9 | 11 | 15 | 5 | 0,074 | 8 | 0,208 |
| 9 | 12 | 14 | 2 | 0,120 | 8 | 0,209 |
| 9 | 13 | 24 | 3 | 0,141 | 7 | 0,276 |
| 9 | 14 | 13 | 0 | - | 0 | - |
| 9 | 15 | 21 | 0 | - | 1 | 0,354 |
| 9 | 16 | 20 | 0 | - | 0 | - |
| 9 | 17 | 10 | 0 | - | 0 | - |
| 9 | 18 | 24 | 0 | - | 1 | 0,392 |
| 9 | 19 | 17 | 0 | - | 2 | 0,480 |
| 9 | 20 | 16 | 0 | - | 2 | 0,406 |
| 9 | 21 | 14 | 0 | - | 0 | 0 |
| 9 | 22 | 19 | 0 | - | 1 | 0,445 |
| 9 | 23 | 17 | 0 | - | 0 | - |