

**MAPEAMENTO DE QTLs EM POPULAÇÕES
COM PEDIGREES COMPLEXOS**

MARIA IMACULADA DE SOUSA SILVA

2008

MARIA IMACULADA DE SOUSA SILVA

**MAPEAMENTO DE QTLs EM POPULAÇÕES COM
PEDIGREES COMPLEXOS**

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

Orientador

Prof. Dr. Júlio Sílvio de Sousa Bueno Filho

**LAVRAS
MINAS GERAIS - BRASIL
2008**

**Ficha Catalográfica Preparada pela Divisão de Processos Técnicos
da Biblioteca Central da UFLA**

Silva, Maria Imaculada de Sousa.

Mapeamento de QTLs em populações com pedigrees complexos/
Maria Imaculada de Sousa Silva - Lavras: UFLA, 2008.
104 p. : il.

Tese(Doutorado) - Universidade Federal de Lavras, 2008.
Orientador: Júlio Sílvio de Sousa Bueno Filho.
Bibliografia.

1. Inferência bayesiana. 2. Saltos reversíveis. 3. Pedigree complexo.
4. Locus controladores de características quantitativas (QTL). 5.
Populações abertas. I. Universidade Federal de Lavras. II. Título.

CDD-519.542

MARIA IMACULADA DE SOUSA SILVA

MAPEAMENTO DE QTLs EM POPULAÇÕES COM
PEDIGREES COMPLEXOS

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

APROVADA em 14 de novembro de 2008

| | |
|---|------|
| Prof. Dr. Antonio Augusto Franco Garcia | USP |
| Prof. Dr. Daniel Furtado Ferreira | UFLA |
| Profa. Dra. Roseli Aparecida Leandro | USP |
| Prof. Dr. Tarcisio de Moraes Gonçalves | UFLA |

Prof. Dr. Júlio Sílvio de Sousa Bueno Filho
UFLA
(Orientador)

LAVRAS
MINAS GERAIS - BRASIL

A Deus, Senhor de tudo
e de toda a minha vida,

Louvo e agradeço

Aos meus filhos, Wellington e Bruno,
ao meu marido, José Diniz,
amores da minha vida.

Dedico

AGRADECIMENTOS

A Deus, por todas as realizações de minha vida e por seu infinito amor e bondade, que me permitiram concluir mais esta etapa.

Aos meus filhos, Wellington e Bruno, e ao meu marido, José Diniz, que são tudo na minha vida, por me darem forças pra continuar, pelo incentivo com muito amor e carinho, e, principalmente, pela compreensão e apoio em cada momento, de dificuldade ou de felicidade.

Ao professor Júlio Sílvio de Sousa Bueno Filho, por sua orientação, amizade e auxílio, imprescindíveis para a realização deste trabalho.

Aos professores membros da banca, pelas contribuições, especialmente ao professor Daniel Furtado Ferreira, pelas conversas e esclarecimentos.

Aos professores do Departamento de Ciências Exatas, por todos os ensinamentos e exemplos de vida e trabalho, ao DEX e à UFLA.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq, pela bolsa de estudos, em grande período de curso.

A toda a minha família, irmãos e irmãs, sobrinhos e sobrinhas, que acreditam e desejam meu sucesso, e que sempre serão especiais.

As minhas grandes amigas, Nádia e Fabrícia, pelo apoio, amizade, ajuda, por todos os momentos compartilhados, de vitórias e dificuldades e pela convivência de irmãs em Lavras.

Às amigas Andréa, Roberta, Luciane e Verônica, pelo apoio e amizade, tornando inesquecíveis os momentos compartilhados, e às amigas Gabriella, Flávia e Franciella, pela convivência jamais esquecida.

Aos colegas do mestrado e do doutorado, que estão em Lavras e os que seguiram outros caminhos, deixando lembranças de grandes alegrias.

Às secretárias e funcionárias do DEX, Josiane, Maria, Vânia, Selminha e Edila, pela dedicação, paciência, amizade e profissionalismo.

A todos aqueles que, de alguma forma, estiveram presentes na minha vida, contribuindo para a conclusão desta etapa.

SUMÁRIO

| | |
|---|----|
| LISTA DE TABELAS | i |
| LISTA DE FIGURAS. | ii |
| RESUMO. | iv |
| ABSTRACT | vi |
| 1 INTRODUÇÃO | 1 |
| 2 REFERENCIAL TEÓRICO | 4 |
| 2.1 Inferência bayesiana | 4 |
| 2.1.1 Amostrador de Gibbs | 7 |
| 2.1.2 Algoritmo de Metropolis-Hastings | 8 |
| 2.1.3 Diagnóstico de convergência | 10 |
| 2.1.4 Comparação entre modelos. | 10 |
| 2.2 Aspectos teóricos do mapeamento genético | 11 |
| 2.2.1 Locos marcadores e QTL | 11 |
| 2.3 Métodos de mapeamento de QTL | 14 |
| 2.3.1 Análise de marcas individuais. | 14 |
| 2.3.2 Mapeamento por intervalo | 15 |
| 2.3.3 Mapeamento por intervalo composto | 16 |
| 2.3.4 Mapeamento de múltiplos intervalos | 17 |
| 2.4 Análise bayesiana de mapeamento por intervalo | 19 |
| 2.5 Modelo QTL para delineamentos com pedigree simples. | 21 |
| 2.6 Modelo QTL em delineamentos com pedigree complexo | 25 |
| 2.6.1 Modelo linear misto | 25 |
| 2.6.2 Distribuições a priori | 27 |
| 2.6.3 Distribuições a posteriori | 28 |
| 2.6.4 Probabilidades de aceitação | 29 |
| 3 METODOLOGIA | 34 |
| 3.1 Simulações | 34 |
| 3.1.1 Simulação 1 (nf=5,nfc=6,nc=2,nQTL=3,dM=10) | 34 |
| 3.1.2 Simulação 2 (nf=5,nfc=10,nc=1,nQTL=2,dM=10) | 36 |

| | | |
|-------|--|----|
| 3.1.3 | Simulação 3 (nf=5,nfc=10,nc=1,nQTL=2,dM=2) | 37 |
| 3.2 | Modelo linear misto e análise bayesiana | 38 |
| 3.2.1 | Distribuições a posteriori | 40 |
| 3.3 | Algoritmos de amostragem | 44 |
| 4 | RESULTADOS E DISCUSSÃO | 50 |
| 4.1 | Resultados da simulação 1 | 50 |
| 4.1.1 | Análise condicional ao número de QTLs. | 56 |
| 4.2 | Resultados da simulação 2 | 60 |
| 4.3 | Resultados da simulação 3 | 64 |
| 5 | CONCLUSÕES | 69 |
| | REFERÊNCIAS BIBLIOGRÁFICAS | 70 |
| | ANEXOS | 77 |

LISTA DE TABELAS

| | | |
|---|--|----|
| 1 | Variâncias alélicas e de dominância simuladas para os 3 QTLs com as respectivas posições no cromossomo. | 36 |
| 2 | Probabilidades a posteriori para o número de QTLs em 100.000 iterações (nf=5, nfc=6, nc=2, nQTL=3, dM=10). | 50 |
| 3 | Estimativas para os parâmetros média geral e variância residual em 100.000 iterações (nf=5, nfc=6, nc=2, nQTL=3, dM=10). (LI limite inferior, LS limite superior, vp valor paramétrico). | 51 |
| 4 | Probabilidades a posteriori para o número de QTL em 9.000 amostras válidas (nf=5, nfc=10, nc=1, nQTL=2, dM=10). | 60 |
| 5 | Estimativas para os parâmetros média geral e variância residual em 9.000 iterações (nf=5, nfc=10, nc=1, nQTL=2, dM=10). (LI limite inferior, LS limite superior, vp valor paramétrico). | 61 |
| 6 | Probabilidades a posteriori para o número de QTLs em 24.0000 amostras válidas (nf=5, nfc=10, nc=1, nQTL=2, dM=2). | 64 |

LISTA DE FIGURAS

| | | |
|----|--|----|
| 1 | Pedigree simulado, consistindo de 65 indivíduos, sendo uma geração de fundadores e duas gerações de descendentes. . . | 35 |
| 2 | Esquema de decisões para o algoritmo com saltos reversíveis. | 48 |
| 3 | Distribuições a posteriori aproximadas para a média geral e variância residual (100.000 iterações, $nf=5$, $nfc=6$, $nc=2$, $nQTL=3$, $dM=10$). | 51 |
| 4 | Intensidade a posteriori de QTL no cromossomo 1, com intervalos de 1cM (100.000 iterações, $nf=5$, $nfc=6$, $nc=2$, $nQTL=3$, $dM=10$). | 52 |
| 5 | Intensidade a posteriori de QTL no cromossomo 2, com intervalos de 1cM (100.000 iterações, $nf=5$, $nfc=6$, $nc=2$, $nQTL=3$, $dM=10$). | 53 |
| 6 | Distribuições a posteriori aproximadas das variâncias aditiva e de dominância do QTL no intervalo de 20 a 30cM (a e b) e no intervalo de 70 a 80cM (c e d) no cromossomo 1 (100.000 iterações, $nf=5$, $nfc=6$, $nc=2$, $nQTL=3$, $dM=10$). | 55 |
| 7 | Distribuições a posteriori aproximadas para a média geral e variância residual (50.000 iterações, $nf=5$, $nfc=6$, $nc=2$, $nQTL=3$, $dM=10$, modelo fixo em 3 QTLs). | 57 |
| 8 | Intensidade a posteriori de QTL nos cromossomos 1 e 2, com intervalos de 1cM (50.000 iterações, $nf=5$, $nfc=6$, $nc=2$, $nQTL=3$, $dM=10$, modelo fixo em 3 QTLs). | 58 |
| 9 | Distribuições a posteriori aproximadas para as variâncias aditiva e de dominância no intervalo de 70 a 80cM no cromossomo 1 (50.000 iterações, $nf=5$, $nfc=6$, $nc=2$, $nQTL=3$, $dM=10$, modelo fixo em 3 QTLs). | 59 |
| 10 | Distribuições a posteriori aproximadas para a média geral e variância residual. (9.000 iterações, $nf=5$, $nfc=10$, $nc=1$, $nQTL=2$, $dM=10$). | 61 |

| | | |
|----|--|-----|
| 11 | Intensidade a posteriori de QTL com intervalos de 1cM no cromossomo 1. (9.000 iterações, nf=5, nfc=10, nc=1, nQTL=2, dM=10). | 62 |
| 12 | Distribuições a posteriori aproximadas das variâncias aditiva e de dominância do QTL no intervalo de 30 a 40cM no cromossomo 1. (9.000 iterações, nf=5, nfc=10, nc=1, nQTL=2, dM=10). | 63 |
| 13 | Distribuições a posteriori aproximadas para a média geral e variância residual (24.000 iterações, nf=5, nfc=10, nc=1, nQTL=2, dM=2). | 65 |
| 14 | Intensidade a posteriori de QTL com intervalos de 1cM no cromossomo 1 (24.000 iterações, nf=5, nfc=10, nc=1, nQTL=2, dM=2). | 66 |
| 15 | Distribuições a posteriori aproximadas das variâncias aditiva e de dominância no intervalo de 30 a 40cM (<i>a</i> e <i>b</i>) e no intervalo de 65 a 85cM (<i>c</i> e <i>d</i>) no cromossomo 1. (24.000 iterações, nf=5, nfc=10, nc=1, nQTL=2, dM=2). | 67 |
| 16 | Distribuições a priori (densidade) e a posteriori (histograma) das variâncias aditiva e de dominância no intervalo de 30 a 40cM (<i>a</i> e <i>b</i>). (9.000 iterações, nf=5, nfc=10, nc=1, nQTL=2, dM=10). | 102 |
| 17 | Distribuições a priori (densidade) e a posteriori (histograma) das variâncias aditiva e de dominância no intervalo de 30 a 40cM (<i>a</i> e <i>b</i>) e de 65 a 85cM (<i>c</i> e <i>d</i>). (24.000 iterações, nf=5, nfc=10, nc=1, nQTL=2, dM=2). | 103 |
| 18 | Distribuições a priori (densidade) e a posteriori (histograma) para a variância residual. (24.000 iterações, nf=5, nfc=10, nc=1, nQTL=2, dM=2). | 104 |

RESUMO

Silva, Maria Imaculada de Sousa. **Mapeamento de QTLs em populações com pedigrees complexos**. 2008. 104 p. Tese (Doutorado em Estatística e Experimentação Agropecuária) - Universidade Federal de Lavras, Lavras, MG. *

Muitos tipos de marcadores moleculares podem ser usados atualmente, em especial para detectar locos controladores de caracteres quantitativos (QTLs). O objetivo usual é identificar posições e efeitos em modelos com número determinado de QTLs. A maior parte dos delineamentos utilizados considera populações originadas de cruzamentos simples (e suas gerações F2 e ou de retrocruzamento). Nestes casos, modelos bialélicos e análise de efeitos fixos para alelos dos QTLs são empregados. Este trabalho teve por objetivo implementar uma análise bayesiana para o mapeamento conjunto de número e efeitos de QTLs mulialélicos, em um contexto de modelos mistos. Um pedigree complexo de duas gerações foi simulado com os respectivos valores genéticos e fenotípicos. Todo o pedigree, fenótipos das marcas e do caráter em estudo foram considerados conhecidos. Um procedimento Monte Carlo via Cadeias de Markov foi implementado para gerar amostras da distribuição a posteriori conjunta, usando o amostrador de Gibbs, quando possível, e o algoritmo Metropolis-Hastings, quando necessário. O método de saltos reversíveis (RJ) foi implementado para permitir que modelos com diferentes números de QTLs fossem amostrados. Os parâmetros desconhecidos nos modelos são a média geral e a variância residual, posições de QTLs, efeitos aditivos de alelos e dominantes de interações alélicas intra-locus, bem como suas respectivas variâncias aditiva e de dominância por loco. Tais parâmetros foram condicionados ao número de QTLs em cada modelo, implicando em modelos com diferentes dimensões. Foram realizadas três simulações. As amostras RJ da primeira, com cinco fundadores e seis filhos por cruzamento, permitiram inferência sobre o número de QTLs, mas a inferência sobre posição e efeitos seria imprecisa. A análise do modelo condicionado no número correto de QTLs (três) não melhorou a inferência sobre a posição e efeitos, em relação ao RJ. Tanto o aumento do número de filhos por cruzamento quanto a maior saturação do mapa resultaram em distribuições a posteriori que permitiram inferência mais precisa, mas os resultados não foram conclusivos. A metodologia mostrou-se promissora *in totum*, mas é preciso desenvolver programas mais rápidos e eficientes para

* **Orientador:** Júlio Sílvio de Sousa Bueno Filho - UFLA.

permitir a análise de pedigrees mais extensos.

Palavras-chave: Inferência bayesiana, saltos reversíveis, pedigree complexo, locos controladores de características quantitativas, populações abertas.

ABSTRACT

Silva, Maria Imaculada de Sousa. **QTL mapping in populations with complex pedigrees.** 2008. 104 p. Thesis (Doctor in Statistics and Agricultural Experimentation) - Universidade Federal de Lavras, Lavras MG.*

Nowadays, many kinds of molecular markers have been used to map loci that controls expression of quantitative traits (QTL). The usual objective is to identify position and effects of QTL, conditioning on a QTL number. Most common QTL designs consider simple pedigree (F2 or backcross populations from a single cross) and biallelic models with fixed QTL effects. This work aimed to implement a bayesian analysis to jointly map number and effects of multiallelic QTL in a mixed model framework. A two-generation complex pedigree with its respective genetic and phenotypic values was simulated. The full pedigree, marker data and phenotypic values were then taken as known observables. A Markov Chain Monte Carlo sampling procedure was turned on to generate samples of the joint posterior, using Gibbs Sampling when possible and Metropolis Hastings sampling when needed. A Reversible Jump (RJ) procedure was implemented. This RJ allowed models with different numbers of QTL to be sampled. Unknown parameters in the models were overall mean and error variance, QTL positions, additive and dominance variance components, allelic additive effects and intra-locus interactions (dominance effects) for each QTL locus. These parameters were conditioned on the number of QTLs in the model, thus implying models with different dimensions. Three simulations were carried out. The first small example with five founders and six individuals per cross has shown that RJ allows inference on QTL number, but a less precise inference on QTL position and effects. Analysis conditioning on correct number of QTLs (three) did not improve inference on QTL position and effects when compared to previous analysis. The second example had more individuals per cross and the third one had a more saturated map. Both examples lead to posterior distributions that would allow better inference, but results are not conclusive. The whole methodology seems promising, but faster and more efficient programs are needed to allow analysing larger pedigrees.

Key-words: Bayesian inference, complex pedigree, outbreeding populations, Quantitative Trait Loci, reversible jump.

* **Adviser:** Júlio Sílvio de Sousa Bueno Filho - UFLA.

1 INTRODUÇÃO

Em estudos de herança genética, é comum associar a variação contínua observada à ação conjunta de efeitos ambientais e de genes ou de grupos de genes de efeito individual indistinguível. Dessa forma, ao se referir à ação desses genes, de forma mais apropriada, refere-se à localização no cromossomo de regiões associadas com a expressão de caracteres quantitativos, utilizando o termo locos controladores de características quantitativas (“quantitative trait loci”, ou QTL). Assim, QTL seria a região cromossômica em que se concentram genes controladores de características quantitativas.

Considerando que a maioria das características de interesse econômico, tais como produção de grãos, peso de frutos, altura de planta e produção de leite, entre outras, possuem variação contínua, existe interesse em determinar posição e efeitos desses QTLs, na tentativa de identificar técnicas de melhoramento capazes, por exemplo, de selecionar indivíduos com fenótipos desejáveis. Ao estudo em que se busca identificar número e posições de QTLs no genoma, quantificando seus efeitos e interações, denomina-se mapeamento de QTL.

A idéia inicial de mapeamento baseava-se na identificação desses QTLs por meio da análise da segregação conjunta entre esses genes e outros genes de manifestação observável, denominados genes marcadores. Posteriormente, com o surgimento dos marcadores moleculares cujos polimorfismos são identificáveis no próprio DNA, tornou-se possível a construção de mapas genéticos mais completos, essenciais para o mapeamento de QTL.

Os métodos mais comuns de mapeamento são a análise de marcas individuais, o mapeamento por intervalo e o mapeamento por intervalo composto. Inicialmente, as técnicas estatísticas mais utilizadas baseavam-se em modelos de regressão e testes de razão de verossimilhanças. Mais recentemente, as técnicas de inferência bayesiana vêm ganhando cada vez mais espaço nas análises de mapeamento, uma vez que as suposições iniciais de delineamento se tornaram mais complexas e o número de parâmetros

a serem estimados aumentou consideravelmente. O avanço dos recursos computacionais permitiu uma maior propagação dos métodos de simulação Monte Carlo via cadeias de Markov (MCMC), tornando possível soluções aproximadas para os problemas de integração múltipla na inferência bayesiana.

Na maioria das abordagens da literatura, número de QTLs é definição prévia do modelo. Nos primeiros trabalhos a utilizarem a inferência bayesiana no mapeamento de QTL (Satagopan et al., 1996), foram ajustados, separadamente, diferentes modelos com números fixos de QTLs e, posteriormente, foi utilizado o fator de Bayes como critério de escolha do modelo mais adequado.

Uma das grandes vantagens da inferência bayesiana, neste contexto, é a possibilidade de estimar o número de QTLs do modelo, conjuntamente com a estimação dos efeitos e posições. Sendo o número de QTLs desconhecido, a dimensão do espaço paramétrico do modelo também é desconhecida e, dessa forma, os métodos de amostragem se tornam mais complicados.

A técnica dos saltos reversíveis, proposta por Green (1995), utiliza os métodos MCMC, permitindo ao algoritmo alternar entre modelos com espaços paramétricos de dimensões diferentes, estimando-se assim o modelo mais adequado aos dados. No contexto de mapeamento de QTL, o algoritmo alterna entre modelos com diferentes números de QTLs estimando número, posições e efeitos de QTLs, simultaneamente.

Em problemas de mapeamento de QTL, os delineamentos mais frequentemente considerados baseiam-se no desequilíbrio de ligação, partindo de cruzamentos controlados entre duas linhagens homozigóticas e contrastantes, sendo possíveis dois alelos por loco. No entanto, esses cruzamentos não ocorrem em populações naturais e nem são sempre viáveis de serem obtidos em qualquer programa de melhoramento. Um dos motivos, por exemplo, é a dificuldade em obter linhagens endogâmicas para determinadas espécies. Os modelos com suposição de QTL bialélico são eficientes

para situações de cruzamentos controlados. No entanto, essa suposição não é razoável para pedigrees complexos, tornando-se necessário estender a modelagem para contemplar a possibilidade de sistemas multialélicos (Yi & Xu, 2001).

Os pedigrees complexos originados de cruzamentos aleatórios entre indivíduos não endogâmicos são, em geral, encontrados em populações naturais e oferecem oportunidades adicionais para detectar QTLs. Embora seja possível simplificar a análise fragmentando-se o pedigree completo em famílias menores, espera-se que o fato de considerar a relação complexa entre os indivíduos possa fornecer maior poder de detecção de QTL (Almasy & Blangero, 1998; George et al., 2000).

A análise bayesiana apresenta grande flexibilidade na capacidade de considerar diversos modelos, pedigrees completos, número desconhecido ou múltiplos QTL, marcadores não informativos ou não genotipados, assim como quaisquer outros parâmetros desconhecidos no problema de mapeamento, que normalmente não seriam possíveis de serem estimados utilizando-se os métodos clássicos de inferência. No entanto, com o aumento da complexidade dos modelos, aumenta também a dificuldade na implementação dos algoritmos, sendo, portanto, necessário buscar técnicas de programação mais eficientes, de forma a minimizar a demanda computacional decorrente dessa complexidade.

Neste trabalho, o objetivo foi a implementação de um método de análise bayesiana de mapeamento de QTL (multialélico em pedigrees complexos), utilizando saltos reversíveis para a inferência marginal ao número de QTLs, em comparação com a análise condicional a esse número.

2 REFERENCIAL TEÓRICO

2.1 Inferência bayesiana

A inferência bayesiana, ao contrário da inferência clássica, leva em conta o conceito de probabilidade subjetiva. Operacionalmente, a maior diferença é tentar medir o grau de incerteza que se tem sobre a ocorrência de um determinado evento do espaço amostral, utilizando distribuições de probabilidades a priori. Assim, a análise bayesiana descreve toda quantidade desconhecida por meio de probabilidades.

Na inferência bayesiana, à incerteza sobre os parâmetros desconhecidos associa-se uma distribuição de probabilidade, enquanto que, na inferência freqüentista, os parâmetros são valores fixos ou constantes, aos quais não se associa qualquer distribuição.

O conhecimento prévio que o pesquisador tem sobre o parâmetro θ é incorporado à análise por meio da densidade a priori $p(\theta)$, a qual deve representar a distribuição de probabilidade desse parâmetro antes da observação dos dados (Box & Tiao, 1992). O parâmetro pode ser um escalar ou um vetor de parâmetros ($\theta = \theta_1, \theta_2, \dots, \theta_p$).

Os parâmetros indexadores das distribuições a priori são chamados de hiperparâmetros, distinguindo-os, assim, dos parâmetros de interesse. Inicialmente, os hiperparâmetros são considerados conhecidos e traduzem a informação que se tem sobre o parâmetro, antes da realização da amostra.

Uma classe importante de densidades a priori são as chamadas prioris conjugadas, as quais conduzem a uma distribuição a posteriori que pertence à mesma classe de distribuições a priori, envolvendo para isso apenas uma mudança no valor dos hiperparâmetros. A tratabilidade analítica das expressões é uma grande vantagem decorrente do uso de prioris conjugadas. No entanto, conforme salientam Gelman et al. (1997), é essencial ter em mente que a distribuição deve adequadamente representar a incerteza a priori sobre os parâmetros, o que nem sempre ocorre com o uso de prioris

conjugadas.

As distribuições a priori podem ser informativas ou não informativas. Quando o pesquisador tem alguma informação prévia sobre o parâmetro em questão, ele pode usar uma priori informativa, descrevendo uma densidade $p(\theta)$ que represente essa informação.

Quando se tem pouca ou nenhuma informação sobre o parâmetro, pode-se usar uma priori não informativa. A idéia é pensar em todos os valores como igualmente prováveis, ou seja, com uma distribuição a priori uniforme. Neste caso, $p(\theta)$ proporcional a uma constante k ($p(\theta) \propto k$) significa que nenhum valor de θ tem preferência. Algumas das maneiras de caracterizar essa informação a priori são detalhadas por Berger (1985), Box & Tiao (1992) e Bernardo & Smith (1994).

A distribuição a priori do parâmetro θ , que depende dos hiperparâmetros ϕ , também pode ser especificada em estágios, sendo denominada, nesses casos, de priori hierárquica, denotada por $p(\theta|\phi)$. Neste caso, os hiperparâmetros não são considerados valores fixos, atribuindo-se também a eles uma distribuição a priori $p(\phi)$, especificando-se assim o segundo estágio na hierarquia. Dessa forma, a distribuição marginal a priori $p(\theta)$ é obtida integrando-se a distribuição $p(\theta, \phi) = p(\theta|\phi) p(\phi)$, em relação a ϕ (Bernardo & Smith, 1994).

A informação sobre θ , resumida probabilisticamente pela distribuição a priori, pode ser aumentada observando-se uma quantidade aleatória y relacionada com θ , sendo o teorema de Bayes a regra de atualização dessa informação. Na análise bayesiana, toda a inferência é feita a partir da distribuição a posteriori $p(\theta|y)$, obtida pelo teorema de Bayes, conforme expressão a seguir:

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{\int p(y|\theta)p(\theta)d\theta}.$$

Em problemas de inferência, convencionou-se dizer que y é o vetor de observações, $p(y|\theta)$ é a função de verossimilhança e $p(\theta)$ é a densidade a priori de θ . O denominador da expressão do teorema de Bayes funciona como uma constante normalizadora, já que não depende de θ . Assim, desconsiderando-se a constante, a distribuição a posteriori é proporcional ao produto entre priori e verossimilhança, ou seja, o teorema de Bayes pode ser reescrito como:

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

No caso de θ ser um vetor de parâmetros, $\theta = (\theta_1, \theta_2, \dots, \theta_p)$, as distribuições marginais dos componentes θ_i , ($i = 1, 2, \dots, p$), a partir das quais as inferências para cada parâmetro serão feitas, podem ser obtidas integrando-se a densidade conjunta a posteriori $p(\theta|y)$ em relação aos demais parâmetros do vetor, ou seja,

$$p(\theta_i|y) = \int \int \dots \int p(\theta_1, \theta_2, \dots, \theta_p|y) d\theta_{-i},$$

sendo $\theta_{-i} = (\theta_1, \dots, \theta_{i-1}, \theta_{i+1}, \dots, \theta_p)$ o vetor θ com sua i -ésima componente removida.

A resolução analítica desta integral é, em geral, impraticável. As alternativas neste caso são métodos de aproximação numérica. Em oposição a esses métodos, outros são baseados em simulação estocástica e fornecem aproximações que serão tanto melhores quanto maior for o número de valores gerados. Os métodos de simulação estocástica consideram as distribuições condicionais completas a posteriori de cada parâmetro para gerar amostras que convergem para a densidade marginal, com o aumento do tamanho dessa amostra (Gelfand & Smith, 1990; Gelfand, 2000).

As distribuições condicionais completas de cada parâmetro, denotadas por $p(\theta_i|\theta_{-i}, y)$, são obtidas considerando-se que, na densidade conjunta, os demais parâmetros θ_{-i} são constantes e, assim, podem ser desconsiderados, levando a uma expressão menos complexa.

Em alguns casos, a expressão da condicional completa tem a forma de uma densidade conhecida e fácil de ser amostrada diretamente. Um método de simulação indicado nesses casos é o amostrador de Gibbs, um processo iterativo que gera valores que convergem para a densidade marginal, sem que se conheça a sua expressão (Gelfand, 2000).

Embora a distribuição condicional completa a posteriori possa ser obtida analiticamente, em determinadas situações, sua expressão não tem a forma de uma densidade conhecida ou fácil de ser amostrada. Em tais situações, outros métodos de simulação são indicados; entre eles, destaca-se o algoritmo de Metropolis-Hastings (Metropolis et al., 1953; Hastings, 1970; Chib & Greenberg, 1995). Os algoritmos do amostrador de Gibbs e do Metropolis-Hastings estão resumidamente descritos a seguir.

2.1.1 Amostrador de Gibbs

O amostrador de Gibbs é um esquema iterativo de amostragem de uma cadeia de Markov. Inicialmente, foi utilizado por Geman & Geman (1984) no contexto de restauração de imagens e, desde então, tem sido muito discutido e utilizado em diversas áreas de pesquisa (Gelfand & Smith, 1990; Gelfand, 2000).

O algoritmo pode ser brevemente descrito, supondo-se que a distribuição de interesse seja $p(\theta|y)$, em que $\theta = (\theta_1, \theta_2, \dots, \theta_p)$. Cada um dos componentes θ_i pode ser um escalar ou um vetor. A distribuição não precisa ser uma distribuição a posteriori e o método pode ser aplicado em qualquer contexto de integração numérica para a obtenção da distribuição conjunta a partir de distribuições condicionais. No caso da inferência bayesiana, no entanto, a distribuição $p(\theta|y)$ corresponde à distribuição a posteriori con-

junta. O interesse aqui é gerar amostras da densidade conjunta, mas, sendo esta geração extremamente complicada, o algoritmo de Gibbs fornece uma forma alternativa de gerações por meio das distribuições condicionais completas. Considerando-se, ainda, que as densidades condicionais completas a posteriori $p(\theta_i|\theta_{-i}, y)$, $i = 1, \dots, p$ estejam disponíveis, o algoritmo é descrito de acordo com os seguintes itens:

- 1 - iniciar o contador de iterações da cadeia $t = 1$ e estabelecer valores iniciais $\theta^0 = (\theta_1^0, \theta_2^0, \dots, \theta_p^0)$;
- 2 - obter um novo valor $\theta^t = (\theta_1^t, \theta_2^t, \dots, \theta_p^t)$ a partir de θ^{t-1} por meio de sucessivas gerações de valores:

$$\begin{aligned} \theta_1^{(t)} &\sim p(\theta_1|\theta_2^{(t-1)}, \theta_3^{(t-1)} \dots, \theta_p^{(t-1)}) \\ \theta_2^{(t)} &\sim p(\theta_2|\theta_1^{(t)}, \theta_3^{(t-1)}, \dots, \theta_p^{(t-1)}) \\ &\vdots \\ \theta_p^{(t)} &\sim p(\theta_p|\theta_1^{(t)}, \theta_2^{(t)} \dots, \theta_{p-1}^{(t)}); \end{aligned}$$

- 3 - mudar o contador t para $t+1$ e volta-se ao passo 2, até que a convergência seja alcançada.

O número de iterações deve ser suficientemente grande para que a seqüência de valores gerados se aproxime da distribuição de equilíbrio, ou seja, da densidade marginal desejada para cada parâmetro, quando se assume que a convergência foi atingida de acordo com Gelfand (2000).

2.1.2 Algoritmo de Metropolis-Hastings

O algoritmo de Metropolis-Hastings é um método de Monte Carlo via cadeias de Markov (MCMC) definido a partir dos trabalhos de Metropolis et al. (1953) e Hastings (1970) (Chib & Greenberg, 1995).

A distribuição condicional completa de interesse, $p(\theta_i|\theta_{-i}, y)$ não tem a forma de uma densidade conhecida, não sendo possível gerar valores

diretamente dessa distribuição. O método consiste em gerar valores candidatos θ_i^* de uma densidade auxiliar $q(\theta_i, \theta_i^*)$ que possa ser amostrada. Os valores candidatos são aceitos com probabilidade dada por $\alpha(\theta_i, \theta_i^*)$, sendo α o mínimo entre 1 e A , ou seja,

$$\alpha(\theta_i, \theta_i^*) = \min\{1, A\} \quad (1)$$

em que:

$$A = \frac{p(\theta_i^*)q(\theta_i, \theta_i^*)}{p(\theta_i)q(\theta_i^*, \theta_i)} \quad (2)$$

A simulação de uma amostra de $p(\theta_i|\theta_{-i}, y)$ pode ser obtida usando o algoritmo Metropolis-Hastings, conforme esquematizado a seguir:

- 1 - iniciar com um valor arbitrário θ_i^j e com o contador de iterações $j = 0$;
- 2 - gerar um valor θ_i^* da distribuição $q(\theta_i^j, .)$ e um valor u de uma uniforme $(0,1)$;
- 3 - calcular $\alpha(\theta_i^j, \theta_i^*)$;
- 4 - se $u \leq \alpha$, aceita-se o novo valor fazendo-se $\theta_i^{j+1} = \theta_i^*$. Caso contrário, faz-se $\theta_i^{j+1} = \theta_i^j$.
- 5 - incrementar o contador j para $j + 1$ e voltar ao passo 2.

De acordo com Chib & Greenberg (1995), várias são as famílias de densidades que podem ser escolhidas para densidade candidata $q(\theta_i, \theta_i^*)$, algumas delas já mencionadas por Hastings (1970). Se a distribuição condicional completa pode ser escrita como $p(\theta_i|\theta_{-i}, y) = \psi(\theta_i)h(\theta_i)$, sendo $h(\theta_i)$ possível de ser amostrada e $\psi(\theta_i)$ é uniformemente limitada, pode-se fazer a distribuição $h(\theta)$ como densidade candidata para gerar as amostras desejadas. Neste caso, a probabilidade de movimento se reduz à seguinte equação:

$$\alpha(\theta_i, \theta_i^*) = \min \left\{ 1, \frac{\psi(\theta_i^*)}{\psi(\theta_i)} \right\} \quad (3)$$

2.1.3 Diagnóstico de convergência

Um dos objetivos da amostragem via métodos MCMC é a convergência para a distribuição marginal dos parâmetros. Assim, torna-se muito importante a monitoração dessa convergência, evitando-se um número excessivo ou insuficiente de iterações no processo de amostragem. Vários trabalhos são dedicados exclusivamente à verificação dessa convergência. Uma avaliação do desempenho dos critérios de convergência mais utilizados foi apresentada por Nogueira et al. (2004).

Um procedimento também muito utilizado é a verificação informal da convergência por meio da análise gráfica das cadeias. Sendo assim, dentre os vários métodos disponíveis na literatura, cabe ao pesquisador interessado definir aquele que melhor atenda aos objetivos de seu estudo.

2.1.4 Comparação entre modelos

Em aplicações práticas, em geral, existem incertezas com relação ao modelo que melhor se ajusta ao conjunto de dados, sendo necessário algum método que auxilie a tomada de decisão em favor do modelo mais plausível. Em inferência bayesiana, um dos critérios usados para esse fim é o Fator de Bayes (Jeffreys, 1961; Kass & Raftery, 1995), sendo este uma alternativa bayesiana para o teste de hipóteses.

A comparação entre os diversos modelos de interesse também pode ser feita iterativamente, considerando o próprio modelo como um parâmetro desconhecido. A idéia é criar uma variável discreta k que seja indicadora do modelo, com probabilidade a priori $p(k)$. A cada k associa-se um vetor de parâmetros θ_k com dimensão n_k , definindo-se, portanto, uma verossimilhança $p(y|\theta_k, k)$ e uma distribuição a priori $p(\theta_k|k)$. A dimensão do vetor

de parâmetros varia de acordo com o modelo, devendo, portanto, ser considerada como um parâmetro a ser estimado. Assim, os métodos MCMC como amostrador de Gibbs ou Metropolis Hastings não podem ser usados para a estimação desse parâmetro, já que estes foram definidos para vetores de parâmetros com dimensão fixa.

Como uma solução para este problema, o método proposto por Green (1995) é um algoritmo MCMC com saltos reversíveis (“Reversible jump”, RJ) que tem a particularidade de permitir a geração de uma cadeia de Markov que salta entre modelos com espaços paramétricos de dimensões diferentes, mantendo as condições necessárias para a convergência do algoritmo (Paulino et al., 2003). De acordo com esse método, supondo que o estado atual da cadeia seja (θ, k) , ou seja, o modelo k possui vetor de parâmetros θ , um novo modelo k' com vetor de parâmetros θ' de dimensão n'_k é proposto, com probabilidade $p_{k,k'}$. Assumindo que $n'_k > n_k$ e que $\theta' = g(\theta, u)$ para uma função determinística g e um vetor aleatório $u \sim q(u)$ com dimensão $n'_k - n_k$, o seguinte algoritmo pode ser utilizado (Ehlers, 2004):

- 1 - propor mudança de (θ, k) para (θ', k') com probabilidade $p_{k,k'}$;
- 2 - gerar $u \sim q(u)$ com dimensão $n'_k - n_k$;
- 3 - aceitar o novo modelo (θ', k') com probabilidade $\min(1, A)$ sendo

$$A = \frac{p(\theta', k')}{p(\theta, k)} \frac{p_{k',k}}{p_{k,k'} q(u)} \left| \frac{\partial(g(\theta, u))}{\partial(\theta, u)} \right|.$$

2.2 Aspectos teóricos do mapeamento genético

2.2.1 Locos marcadores e QTL

No mapeamento genético em geral, existe grande interesse em detectar regiões cromossômicas responsáveis por alguma variação observada em características quantitativas, assim como em determinar a forma de ação dos genes ou grupos de genes associados a essas regiões. Para se referir a essas regiões cromossômicas, o termo locos controladores de características

quantitativas (“quantitative trait loci”, QTL) foi proposto por Geldermann (1975) e, desde então, tem sido comumente empregado em estudos em que se busca localizar ou identificar esses genes.

Considerando que o efeito dos QTLs não é observável, a idéia inicial, baseada na ligação genética, é identificá-los por meio de genes de manifestação observável, ou seja, de características qualitativas, localizados no mesmo cromossomo, em posições próximas ou ligados. A identificação dos QTLs é feita por meio da análise da segregação conjunta entre os QTLs e esses genes, denominados de locos marcadores.

Se um determinado loco marcador, de fácil observação fenotípica, estiver fisicamente ligado, a pequena distância, a um gene que controla um caráter de interesse, a seleção deste marcador resulta na provável seleção indireta do gene de interesse, exceto quando ocorrerem recombinações (Ferreira & Grattapaglia, 1995).

Após a identificação de um determinado número de genes marcadores para uma espécie de interesse, os mapas genéticos podem ser construídos, mostrando a ordem dos locos no cromossomo e a distância relativa entre eles, em termos de frequência de recombinação. A recombinação ocorre quando existe troca aleatória de material genético durante a meiose, ou seja, dois genes localizados em um mesmo cromossomo não segregam juntos, originando, assim, os gametas recombinantes.

As frequências de recombinação não são aditivas e, portanto, não são apropriadas como medidas de distância (Lynch & Walsh, 1998). As funções de mapeamento foram desenvolvidas para transformar a frequência de recombinação em uma unidade de distância entre locos que seja aditiva, dada em Morgans (M) ou em centésimos de Morgan, centimorgans (cM). Uma unidade de Morgan (1M) corresponde a um segmento cromossômico ao longo do qual se espera que ocorra uma recombinação por gameta por geração.

As funções de mapeamento mais comumente usadas são a função de

Haldane (Haldane, 1919) e a função de Kosambi (Kosambi, 1944).

Inicialmente, os genes marcadores eram todos de características morfológicas. Um dos primeiros trabalhos a utilizar esses marcadores foi apresentado por Sax (1923), quando foi detectada associação entre genes controlando um caráter qualitativo (cor) e genes controlando um caráter quantitativo (peso) em grãos de feijão.

Apesar de os marcadores morfológicos terem contribuído de forma significativa para o desenvolvimento teórico da análise da ligação gênica e para a construção das primeiras versões dos mapas genéticos, a probabilidade de se encontrar associações positivas entre marcadores e QTL era reduzida, devido ao pequeno número de genes marcadores distintos para a maioria das espécies de importância econômica (Ferreira & Grattapaglia, 1995). Assim, por um período de aproximadamente 60 anos, em razão desse reduzido número de genes marcadores disponíveis, os estudos envolvendo a construção de mapas genéticos a partir de marcadores morfológicos não foram muito numerosos, embora muitos deles tenham apresentado contribuições relevantes ao assunto (Elston & Stewart, 1973; Soller et al., 1976; Soller & Genizi, 1978).

O surgimento dos marcadores moleculares que permitem obter informações sobre polimorfismos diretamente no DNA, a partir de 1980, possibilitou um considerável avanço na construção de mapas genéticos mais completos e, conseqüentemente, mais precisos em detectar QTL a partir dessas informações (Tanksley et al., 1982; Weller, 1986; Edwards et al., 1987; Weller et al., 1988).

O aumento do número de locos marcadores disponíveis para se realizar o mapeamento genético exigiu esforços para o desenvolvimento de bases teóricas fundamentais e de métodos de análise que levassem em conta esse aumento de informação (Lander & Botstein, 1989; Haley & Knott, 1992; Jansen, 1992, 1993; Zeng, 1993, 1994).

As reduções em custos computacionais também foram decisivas para

avanços em termos de execução das análises decorrentes desses estudos em que se busca mapear QTL. Mapear QTLs significa determinar seu número e posições no genoma, quantificar seus efeitos, tais como efeitos aditivos e de dominância, interações QTL ambiente e entre locos.

As populações mais frequentemente utilizadas no mapeamento de QTL são aquelas obtidas a partir do cruzamento entre duas linhagens homozigóticas fenotipicamente contrastantes, tais como as gerações segregantes F2 e retrocruzamentos (Lynch & Walsch, 1998).

2.3 Métodos de mapeamento de QTL

2.3.1 Análise de marcas individuais

O mapeamento de QTL pode basear-se na avaliação de uma, duas ou mais marcas simultaneamente, originando, assim, os diversos métodos para mapear QTL. O mais simples desses métodos é chamado análise de marcas individuais. De acordo com este método, independentemente da quantidade de marcadores disponíveis, a associação entre o QTL e um único marcador é testada separadamente (Elston & Stewart, 1973; Stewart & Elston 1973; Soller & Genizi, 1978; Weller, 1986). Assim, a análise de um cromossomo com m marcadores exige m testes de marcas individuais.

A identificação de um QTL ligado a um marcador baseia-se na comparação entre as médias das classes genotípicas dos marcadores. Havendo diferença significativa entre essas médias, existe possibilidade de um QTL associado ao marcador em questão.

Os testes estatísticos utilizados na comparação das médias são o teste t (Soller et al., 1976), o teste de razão de verossimilhança (Weller, 1986), a análise de variância e a regressão linear, simples ou múltipla (Liu, 1998; Lynch & Walsh, 1998).

A análise de marcas individuais é muito utilizada quando se deseja apenas a detecção de um QTL ligado à marca, ou quando o objetivo é descartar marcas não informativas, como um passo preliminar para uma

análise mais complicada (Lynch & Walsh, 1998).

Uma desvantagem desse método, segundo Liu (1998), é que o efeito de um QTL fica confundido com sua posição, que não pode ser precisamente determinada. Além disso, o método não é muito poderoso e exige um grande número de indivíduos amostrados para o teste.

2.3.2 Mapeamento por intervalo

O mapeamento por intervalo considera um par de marcas adjacentes e tenta inferir sobre a existência de um QTL em qualquer posição entre essas duas marcas.

O método do mapeamento por intervalo foi proposto por Lander & Botstein (1989) como uma alternativa para eliminar alguns dos principais problemas apresentados pela análise de marcas individuais.

Esse método permite obter estimativas mais precisas de efeito e posição do QTL, além de aumentar o poder de detecção em relação à análise de marcas individuais. Outra vantagem desse método é que o número de indivíduos necessários para a análise é menor que no método de marcas individuais.

Um modelo linear é usado para associar o fenótipo ao genótipo do QTL, e as estimativas dos parâmetros são obtidas pelo método da máxima verossimilhança. O teste de razão de verossimilhança, ou LOD Score, é usado para verificar a existência de ligação entre os marcadores e o QTL, de forma similar à análise de marcas individuais (Lynch & Walsh, 1998).

Assumindo uma característica quantitativa y de uma população de retrocruzamento, o modelo de regressão linear proposto por Lander e Botstein (1989) para testar a presença de um QTL entre os marcadores j e $j+1$, é dado por:

$$y_i = \beta_0 + \beta x_i + \epsilon_i;$$

em que β_0 é uma constante, β é o efeito do QTL, x_i é uma variável indica-

dora do genótipo do QTL, com probabilidade dependendo do genótipo dos marcadores j e $j+1$, e ϵ_i é o erro aleatório, $i = 1, 2, \dots, n$.

O modelo apresentado é um modelo de misturas, pelo fato de não se conhecer o genótipo do QTL. Assim, explorando as propriedades desse modelo, Lander e Botstein (1989) propuseram um teste de razão de verossimilhanças baseado na hipótese $H_0 : \beta = 0$ e $H_1 : \beta \neq 0$, assumindo que o QTL esteja localizado no intervalo sob consideração. A hipótese H_0 refere-se à ausência de um QTL ligado ao intervalo marcador.

Um dos problemas apresentados pelo mapeamento por intervalo é a detecção de falsos positivos, ou QTLs “fantasmas” (Halley & Knott, 1992; Martinez & Curnow, 1992). Esses QTLs são detectados quando os testes estatísticos indicam a presença de um QTL no intervalo analisado, quando, na verdade, está localizado em intervalos adjacentes, não considerados pela análise (Lynch & Walsh, 1998). Em geral, quando existem dois ou mais QTLs sobre um cromossomo, a posição estimada de um QTL é provavelmente viesada, devido à influência do outro QTL no processo de estimação (Knott & Halley, 1992).

2.3.3 Mapeamento por intervalo composto

Idealmente, quando se testa a presença de um QTL em um intervalo, deseja-se que este teste seja independente dos efeitos de possíveis QTLs em outras regiões do cromossomo (Zeng, 1994). Este teste pode ser construído utilizando-se uma combinação do mapeamento por intervalo com o modelo de regressão múltipla, cujas propriedades também foram analisadas por Rodolphe & Lefort (1993). Com este enfoque, os métodos do mapeamento por intervalo composto foram propostos, independentemente, por Jansen (1992, 1993) e Zeng (1993, 1994).

De acordo com este método, os efeitos dos QTLs fora do intervalo considerado são inseridos no modelo sob a forma de cofatores, sendo, portanto, eliminados do resíduo. Como resultado, tem-se um considerável au-

mento de precisão no mapeamento e a detecção de QTLs “fantasmas” é reduzida (Jansen, 1993; Zeng, 1994).

O modelo de regressão adotado para esse método, de acordo com Zeng (1994), é dado por:

$$y_i = \beta_0 + \beta_j x_{ji} + \sum_{k \neq j, j+1}^M \beta_k x_{ki} + \epsilon_i;$$

em que y_i é o valor observado da característica para o i -ésimo indivíduo, β_j é o efeito do QTL supostamente localizado entre os marcadores j e $j + 1$, x_{ji} é uma variável indicadora do genótipo do QTL, β_k são os coeficientes de regressão parciais para o caráter na marca k , x_{ki} são variáveis indicativas do genótipo para a marca k e ϵ_i é o erro aleatório. De forma similar aos métodos anteriores, as estimativas dos parâmetros podem ser obtidas pelo método da máxima verossimilhança.

Uma extensão do método de mapeamento por intervalo composto foi apresentada por Jiang & Zeng (1995) para mapeamento de QTL, considerando mais que uma característica em um mesmo modelo. Segundo esses autores, um modelo considerando a estrutura de correlação entre os múltiplos caracteres teria maior poder de detecção de QTL e maior precisão de mapeamento em geral, quando comparado com as análises feitas separadamente.

2.3.4 Mapeamento de múltiplos intervalos

O mapeamento de múltiplos intervalos, proposto por Kao et al. (1999), é a estratégia de mapear vários QTLs simultaneamente, em intervalos distintos, obtendo-se um aumento na precisão e no poder de sua detecção.

O mapeamento de múltiplos intervalos permite a estimação dos parâmetros já existentes nos métodos anteriores, como posição e efeitos do QTL, além da possibilidade de estimar o número de QTLs no genoma e os

efeitos de interação entre eles, denominados de efeitos epistáticos (Kao et al., 1999).

No mapeamento por intervalos múltiplos, assume-se que exista um número l de QTLs controlando um caráter quantitativo e que estes estejam localizados em l posições diferentes, flanqueados por marcadores distintos. O genótipo de cada QTL no intervalo considerado é desconhecido, mas sua distribuição pode ser inferida a partir dos marcadores flanqueadores, com base na frequência de recombinação entre eles.

O modelo de regressão que associa a característica observada y aos l QTLs é dado por:

$$y_i = \mu + \sum_{j=1}^l \alpha_j x_{ij} + \sum_{j \neq k}^l \beta_{jk} (a_{jk} x_{ij} x_{ik}) + \epsilon_i;$$

em que μ é uma constante, α_j é o efeito do j -ésimo QTL, x_{ij} é uma variável indicativa do genótipo do j -ésimo QTL, podendo assumir os valores $\frac{1}{2}$ e $-\frac{1}{2}$ considerando uma população de retrocruzamento, a_{jk} é o efeito de epistasia entre os QTLs j e k , β_{jk} é uma variável indicadora para o efeito de epistasia entre os QTLs j e k , com valor 1, se existir efeito de epistasia e 0, caso contrário e ϵ_i é o erro aleatório.

O modelo apresentado para o mapeamento por intervalos múltiplos é uma mistura de distribuições, geralmente normais, sendo um dos principais problemas desse método referente à estimação dos parâmetros dessa mistura, pois, o número de parâmetros aumenta consideravelmente com o aumento do número de QTLs. Dessa forma, a obtenção dessas estimativas pelo método da máxima verossimilhança torna-se impraticável para um número elevado de QTLs.

Uma importante questão a ser considerada para a implementação desse método é a seleção prévia das marcas que poderão fazer parte do modelo. A seleção dos marcadores pode ser feita utilizando-se o processo de seleção de covariáveis “stepwise”, conforme sugerido por Kao et al. (1999),

ou um outro processo similar, para identificar aqueles marcadores que sejam realmente informativos. Outra alternativa seria descartar as marcas não informativas por meio da análise de marcas individuais, utilizando-se procedimentos como a correção de Bonferroni ou a FDR (false discovery rate) para definir o nível de significância (Silva & Vencovsky, 2002).

2.4 Análise bayesiana de mapeamento por intervalo

A análise de dados de mapeamento por intervalo, utilizando-se a abordagem bayesiana, tem se mostrado uma eficiente alternativa na detecção de múltiplos QTL e estimação de seus efeitos, de acordo com vários trabalhos utilizando esse enfoque (Satagopan & Yandell, 1996; Satagopan et al., 1996; Heath, 1997; Uimari & Hoeschele, 1997; Jansen et al., 1998; Stephens & Fisch, 1998; Sillanpaa & Arjas, 1998, 1999; Bink & Arendonk, 1999; Bink et al., 2000; Yi & Xu, 2001; Yi, 2004).

A análise bayesiana apresenta conveniência e flexibilidade no mapeamento de múltiplos QTLs, além de permitir que sejam estimados simultaneamente todos os parâmetros desconhecidos no problema do mapeamento, como efeitos, posição, número de QTLs e, ainda, genótipos de marcador e QTLs (Yi & Xu, 2001). No entanto, a demanda computacional é consideravelmente maior neste tipo de análise.

Uma das alternativas utilizadas para estimar o número total de QTLs em todo o genoma é o ajuste independente de vários modelos com diferentes números de QTLs. Após serem ajustados esses modelos, um critério de seleção, como, por exemplo, o fator de Bayes, é utilizado na escolha do modelo mais adequado aos dados, estimando-se, por consequência, o número de QTLs a influenciar a característica considerada (Satagopan et al., 1996; Sen & Churchill, 2001). Gofinett & Gerber (2000) propuseram algumas modificações no critério de Akaike e sugeriram esse método como critério de escolha do modelo com número mais adequado de QTLs.

O método MCMC com saltos reversíveis (RJ) também pode ser

usado como critério de escolha do modelo mais adequado (Satagopan & Yandell, 1996; Heath, 1997) e, por consequência, de estimação do número de QTLs que, nesse caso, é considerado como um parâmetro a ser estimado. A cada modelo associa-se um número de QTLs e a cadeia de Markov salta entre esses modelos, cujos espaços paramétricos possuem diferentes dimensões. A distribuição da proporção de vezes que a cadeia passa em um determinado modelo é uma estimativa Monte Carlo da distribuição a posteriori do número de QTLs afetando a característica (Sorensen & Gianola, 2002). Essa metodologia permite inferir sobre os demais parâmetros de forma marginal ao número de QTLs, ao contrário de todas as demais técnicas, que inferem sobre posições e efeitos, de forma condicional ao número de QTLs. A aplicabilidade desse método, que permite ao processo iterativo transitar entre os modelos com diferentes números de QTLs a cada iteração, tem sido verificada em vários trabalhos que o utilizaram como processo de estimação do número de QTLs (Stephens & Fisch, 1998; Sillanpää & Arjas, 1998, 1999; Yi & Xu, 2001).

As baixas taxas de aceitação e as dificuldades em se obter uma rápida convergência são os principais problemas do algoritmo com saltos reversíveis (Yi, 2004). Em estudos mais recentes de mapeamento, são propostas novas alternativas de inferência conjunta sobre número e efeito de QTLs tratando cada marcador como um QTL potencial (Xu, 2003), ou supondo que para cada intervalo marcador existe um QTL associado, sendo o genoma supersaturado de QTLs (Yi, 2004; Wang et al., 2005).

No mapeamento de QTLs em geral, destacam-se duas estratégias distintas. Na primeira delas, o delineamento experimental inicia-se com o cruzamento entre duas linhagens endogâmicas, obtendo-se um pedigree simples, com o objetivo de explorar o desequilíbrio de ligação esperado nas populações derivadas desse cruzamento. A segunda baseia-se na análise de um pedigree com muitos indivíduos resultantes de cruzamentos livres. Neste caso, procura-se explorar a natureza da relação genética entre todos

os indivíduos do pedigree complexo. Note-se que no pedigree complexo são admitidos múltiplos alelos por loco, enquanto no pedigree simples apenas dois alelos são esperados.

A grande maioria dos métodos de análise bayesiana, para mapeamento por intervalo, assume a primeira estratégia, ou seja, assume linhagens homozigóticas fenotipicamente contrastantes, conduzindo, assim, a um modelo com QTL bialélico (Satagopan & Yandell, 1996; Satagopan et al., 1996; Sillanpää & Arjas, 1998, 1999). Yi & Xu (2001) afirmam que, embora essa suposição seja razoável para cruzamentos específicos, o mesmo não ocorre para cruzamentos com pedigrees complexos, o que torna necessário estender o método para sistemas multialélicos.

Com um delineamento envolvendo múltiplos cruzamentos espera-se obter maior poder de detecção de QTL do que em delineamentos envolvendo um único cruzamento (George et al., 2000; Yi & Xu, 2001). Segundo esses autores, existe uma necessidade crescente do desenvolvimento de técnicas estatísticas capazes de mapear QTL em dados originados de pedigrees complexos. Em tais casos, destaca-se a importância da inferência bayesiana, uma vez que seria difícil desenvolver o algoritmo usando o método de estimação da máxima verossimilhança. Uma modelagem específica se apresenta adequada para cada tipo de população, de acordo com as particularidades dos cruzamentos considerados.

2.5 Modelo QTL para delineamentos com pedigree simples

Considerando uma característica quantitativa controlada por um número l de QTLs, o modelo utilizado para descrever os dados y (Satagopan et al., 1996) de uma população derivada de cruzamentos controlados entre duas linhagens endogâmicas é dado por:

$$y_i = \mu + \sum_{j=1}^l \alpha_j Q_{ij} + \sum_{j=1}^l \delta_j (1 - |Q_{ij}|) + \epsilon_i; \quad (4)$$

sendo y_i o valor fenotípico observado para o i -ésimo indivíduo, $Q_j = \{Q_{ij}\}_{j=1}^l$ o indicador do genótipo do QTL para o i -ésimo indivíduo, ϵ_i é o erro aleatório com distribuição normal com média zero e variância σ^2 e $\alpha = \{\alpha_j\}_{j=1}^l$ e $\delta = \{\delta_j\}_{j=1}^l$ são, respectivamente, os efeitos aditivos e de dominância dos l QTLs. Os parâmetros de interesse são as posições dos QTLs, $\lambda = \{\lambda_j\}_{j=1}^l$ e o vetor $\theta = (\mu, \alpha, \delta, \sigma^2)$.

Os possíveis efeitos de epistasia podem ser contemplados pelo modelo, incorporando-se a ele termos de interação entre os QTLs.

Satagopan et al., (1996) assumem apenas um grupo de ligação com marcadores ordenados $\{1, 2, \dots, m\}$ e distância conhecida, embora mais grupos de ligação possam ser considerados (Sillanpää & Arjas, 1998).

A distribuição do genótipo dos QTLs, $p(Q|\lambda)$ é obtida, dada a sua localização, em termos de fração de recombinação entre os marcadores flancadores do QTL no intervalo considerado, com $Q = \{Q_i\}_{i=1}^n$ denotando os genótipos dos QTLs para todos os n indivíduos. No caso de indivíduos de uma população F_2 , Satagopan et al. (1996) associam 1 para o genótipo QQ , -1 para o genótipo qq e 0 para o genótipo Qq .

A função de verossimilhança, tendo sido o erro especificado como tendo distribuição normal, também resulta em densidade normal, com média e variância específicas.

As densidades a priori de todos os parâmetros, $p(\lambda, \theta)$, são especificadas de acordo com o conhecimento prévio do pesquisador sobre cada parâmetro.

Uma distribuição a priori comumente utilizada para o parâmetro λ é uma uniforme em um intervalo determinado de acordo com o comprimento do cromossomo. A distribuição a priori frequentemente assumida para o número de QTLs l é uma Poisson truncada em um valor máximo prefixado. Para os demais parâmetros, Satagopan et al., (1996) sugerem prioris difusas ou pouco informativas.

A inferência bayesiana baseia-se em amostras da distribuição con-

junta a posteriori de todos os parâmetros, a qual é obtida após terem sido definidas a função de verossimilhança e todas as prioris, de acordo com a expressão:

$$p(\lambda, \theta, Q|y) \propto p(y|\theta, Q)p(Q|\lambda)p(\lambda, \theta);$$

em que $p(y|\theta, Q) = \prod_{i=1}^n p(y_i|\theta, Q_i)$ é a distribuição dos dados (verossimilhança) e $p(Q|\lambda) = \prod_{i=1}^n p(Q_i|\lambda)$ é a probabilidade do genótipo dos QTLs dadas suas posições.

As distribuições condicionais completas a posteriori de cada parâmetro, a partir das quais as amostras da distribuição de interesse serão geradas, são obtidas da distribuição conjunta.

A estimação do vetor de parâmetros θ pode ser realizada utilizando-se o algoritmo do amostrador de Gibbs ou do Metropolis-Hastings, dependendo da forma da distribuição condicional completa em cada caso.

Na literatura, existem diversas propostas para a estimação do número de QTLs. Entre elas estão a análise condicional e a análise marginal ao número de QTLs em que, este é tratado como um parâmetro desconhecido no modelo. Nas estratégias mais comuns de inferência condicional ao número de QTLs, o fator de Bayes pode ser usado como critério de escolha do modelo (Satagopan et al., 1996). O número de QTLs estimado é aquele considerado no modelo indicado pelo fator de Bayes.

A abordagem considerando o número de QTLs como desconhecido, e portanto um parâmetro a ser estimado no modelo, utiliza, para a estimação desse parâmetro, os métodos MCMC com saltos reversíveis. A eficiência desse método, por vezes denominado algoritmo Metropolis-Hastings-Green (Sillanpää & Arjas, 1998), tem sido comprovada nos vários trabalhos que se propõem a verificar ou a utilizar suas propriedades no mapeamento de QTLs (Satagopan & Yandell, 1996; Heath, 1997; Stephens & Fisch, 1998; Sillanpää & Arjas, 1998, 1999; Yi & Xu, 2001).

As posições dos QTLs podem ser estimadas normalmente por meio

das distribuições condicionais completas desse parâmetro, dependendo da sua forma, usando amostrador de Gibbs ou Metropolis-Hastings. As estimativas a posteriori podem ser resumidas utilizando-se o método da intensidade de QTL a posteriori, proposto por Sillanpää & Arjas (1998). De acordo com esses autores, o principal interesse da análise está no número de QTLs e nas suas posições, podendo o resultado da análise ser expresso em termos da intensidade estimada de QTL.

Na prática, o método da intensidade de QTL consiste em dividir o cromossomo em I intervalos $\Delta_1, \Delta_2, \dots, \Delta_I$ de igual comprimento. O número de ciclos (iterações) no processo MCMC é denotado por N_c e o número de QTLs no modelo na k -ésima iteração é denotado por N_{Qk} .

A intensidade de QTL a posteriori aproximada, no intervalo Δ_j , é obtida via simulação Monte Carlo pela expressão:

$$h(\hat{\lambda}_j) = \frac{1}{\|\Delta_j\|} \left(\frac{1}{N_c} \sum_{k=1}^{N_c} \sum_{q=1}^{N_{Qk}} I(l_q^{(k)} \in \Delta_j) \right); \quad (5)$$

em que $\sum_{q=1}^{N_{Qk}} I(l_q^{(k)} \in \Delta_j)$ é o número de QTLs no intervalo Δ_j no ciclo k das iterações. O produto $h(\hat{\lambda}_j)\|\Delta_j\|$ dá uma aproximação do número esperado a posteriori de QTLs no intervalo Δ_j (Sillanpää & Arjas, 1998).

As densidades a posteriori das variâncias aditiva e de dominância podem ser estimadas de acordo com o método sugerido por Sillanpää & Arjas (1999). Considerando a função de distribuição acumulada associada com a variância aditiva $f_a(\Delta_j)$ e com a variância de dominância $f_d(\Delta_j)$ de um QTL localizado em um pequeno intervalo (Δ_j) do cromossomo, as médias das variâncias amostradas em (Δ_j) são usadas para obter as expressões:

$$f_a(\Delta_j) = \frac{(\sum_{k=1}^{N_c} \sum_{q=1}^{N_{Qk}} \sigma_{a_q}^2 I(\lambda_q^{(k)} \in \Delta_j))}{(\sum_{k=1}^{N_c} \sum_{q=1}^{N_{Qk}} I(\lambda_q^{(k)} \in \Delta_j))}; \quad (6)$$

e

$$f_d(\Delta_j) = \frac{(\sum_{k=1}^{N_c} \sum_{q=1}^{N_{Q^k}} \sigma_{d_q}^2 I(\lambda_q^{(k)} \in \Delta_j))}{(\sum_{k=1}^{N_c} \sum_{q=1}^{N_{Q^k}} I(\lambda_q^{(k)} \in \Delta_j))}. \quad (7)$$

É importante destacar que $f_a(\Delta_j)$ e $f_d(\Delta_j)$ são consideradas boas estimativas somente quando um número suficiente de amostras está contido no intervalo (Δ_j) .

2.6 Modelo QTL em delineamentos com pedigree complexo

2.6.1 Modelo linear misto

A modelagem bayesiana para dados de mapeamento com QTL multialélico pode ser exemplificada considerando dados de uma população consistindo de n indivíduos, dos quais m são fundadores e $(n - m)$ são não fundadores. Os indivíduos fundadores são aqueles que não possuem os pais incluídos no pedigree. Supõe-se que a população seja mapeada para um número k de marcadores e que a característica em questão seja controlada por um número desconhecido l de QTLs. O modelo linear misto a seguir foi apresentado por Yi & Xu (2001), para representar a variação nos dados.

$$y = Xb + \sum_{j=1}^l (u_j^m + u_j^p + v_j) + \mathbf{e}. \quad (8)$$

Em tal modelo, y é o vetor de n valores fenotípicos observados, b é um vetor de efeitos fixos, X é a matriz (conhecida) de delineamento do efeito fixo b , l é o número total de QTLs, u_j^m e u_j^p são vetores de n efeitos alélicos maternos e paternos do j -ésimo QTL, v_j é um vetor de n efeitos de dominância do j -ésimo QTL, \mathbf{e} é um vetor de resíduos (efeitos ambientais) supostos de distribuição normal com média zero e variância σ_e^2 .

Os efeitos dos QTLs, para todos os indivíduos, podem ser expressos como função linear dos efeitos alélicos dos fundadores e suas interações, ou seja, $u_j^m = Z_j^m a_j$; $u_j^p = Z_j^p a_j$ e $v_j = W_j d_j$, sendo a_j um vetor de

$2m$ efeitos alélicos dos fundadores e d_j um vetor (de ordem $m(2m + 1)$) de efeitos de interação entre todos os possíveis pares dos $2m$ alelos dos fundadores para o j -ésimo QTL. Z_j^m e Z_j^p são as matrizes desconhecidas de delineamento dos efeitos alélicos aditivos (de ordem $n \times 2m$), nas quais os alelos do j -ésimo QTL para cada indivíduo são relacionados com os alelos dos fundadores. W_j (de ordem $n \times m(2m + 1)$) é uma matriz desconhecida de delineamento dos efeitos de dominância, que é completamente determinada com o conhecimento das matrizes de herança alélica Z_j^m e Z_j^p .

Considerando as reparametrizações mencionadas, o modelo linear (8) pode ser reescrito de acordo com a expressão a seguir (Yi & Xu, 2001).

$$y = Xb + \sum_{j=1}^l (Z_j^m + Z_j^p)a_j + \sum_{j=1}^l W_j d_j + e. \quad (9)$$

As distribuições das matrizes de herança alélica são funções da informação contida na matriz M de marcadores, e da posição do j -ésimo QTL no cromossomo. A matriz de marcadores inclui a posição dos marcadores no cromossomo e os respectivos genótipos observados. Em alguns casos, a matriz de marcadores M possui valores desconhecidos (não observados), sendo necessário incluir na modelagem um algoritmo para a estimação desses valores perdidos, a partir da informação dos demais marcadores e do pedigree, como ocorre nos trabalhos de Heath (1997), Uimari & Hoeschele (1997) e Bink & Arendonk (1999). A descrição de um algoritmo para a estimação desses genótipos perdidos pode ser encontrada em Jiang & Zeng (1997).

Os valores observáveis no modelo são o vetor de valores fenotípicos y , o vetor X e a matriz de marcadores M .

Os parâmetros de interesse são o número total de QTLs l , as posições $\lambda = \{\lambda_j\}_{j=1}^l$ dos respectivos QTLs no cromossomo, os efeitos alélicos $a = \{a_j\}_{j=1}^l$ de todos os QTLs, os efeitos de dominância $d = \{d_j\}_{j=1}^l$, a variância residual σ_e^2 e as matrizes de herança alélica $Z^m = \{Z_j^m\}_{j=1}^l$ e $Z^p = \{Z_j^p\}_{j=1}^l$.

Os efeitos aditivos e de dominância dos QTLs podem ser considerados como aleatórios, tendo-se, assim, interesse em seus componentes de variância, sendo eles, respectivamente, a variância aditiva $V_a = \{\sigma_{a_j}^2\}_{j=1}^l$ e a variância de dominância $V_d = \{\sigma_{d_j}^2\}_{j=1}^l$. O vetor de parâmetros de interesse θ é, então, representado por: $\theta = \{l, \lambda, a, d, V_a, V_d, b, Z^m, Z^p, \sigma_e^2\}$.

Para fins práticos, em geral, não interessa a estimação específica de cada efeito individual de alelos e interações alélicas, mas a boa estimativa dos vetores de variâncias V_a e V_d além de σ_e^2 . Dessa forma, especifica-se completamente a distribuição de a e d e determina-se a importância relativa da região cromossômica na expressão do caráter.

Supondo que o valor fenotípico seja uma característica normalmente distribuída, a função de verossimilhança é dada pela expressão:

$$p(y|\theta) \propto (\sigma_e^2)^{-\frac{n}{2}} \exp \left\{ -\frac{1}{2\sigma_e^2} R^\top R \right\}, \quad (10)$$

sendo $R = y - Xb - \sum_{j=1}^l (Z_j^m + Z_j^p) a_j - \sum_{j=1}^l W_j d_j$.

O modelo apresentado é analisado por meio dos métodos bayesianos, os quais se baseiam na especificação de distribuições a priori e na obtenção da distribuição a posteriori conjunta. As distribuições a priori para cada parâmetro são apresentadas a seguir.

2.6.2 Distribuições a priori

A escolha mais frequente de distribuição a priori para o número de QTLs l é a distribuição Poisson truncada em um valor prefixado l_{maximo} e com média l_0 (Satagopan et al., 1996; Yi & Xu, 2001). A posição do j -ésimo QTL λ_j é proporcional ao comprimento do cromossomo. Assim, a probabilidade a priori de cada QTL pertencer a uma posição qualquer do cromossomo é a mesma, ou seja, a distribuição atribuída é uniforme.

Uma distribuição a priori indicada para os efeitos aleatórios do j -

ésimo QTL, a_j e d_j , é uma normal com média zero e variância que depende das variâncias aditiva e de dominância e do coeficiente de endogamia dos fundadores (Yi & Xu, 2001). De acordo com esses autores, as distribuições a priori dos efeitos aditivos a_j são especificadas por $p(a_j) = \prod_{k=1}^m p(a_k^j)$, em que $a_k^j = (a_{k(1)}^j, a_{k(2)}^j) \sim N(0, \text{var}(a_k^j))$. O vetor a_k^j especifica o par de efeitos alélicos do k -ésimo fundador para o j -ésimo QTL, cuja variância é dada por

$$\text{var}(a_k^j) = \sigma_{a_j}^2 \begin{bmatrix} 1 & f_k \\ f_k & 1 \end{bmatrix}$$

sendo f_k coeficiente de endogamia do k -ésimo fundador. Se o coeficiente de endogamia for considerado igual a zero, a variância se reduz a $\text{var}(a_k^j) = \sigma_{a_j}^2 I$, sendo I a matriz identidade.

Para os efeitos de dominância d_j do j -ésimo QTL, $p(d_j) = \prod_{k=1}^m \prod_{k'=k}^m p(d_{kk'}^j)$, sendo $d_{kk'}^j \sim N(0, \text{var}(d_{kk'}^j))$. O vetor $d_{kk'}^j$ inclui todos os possíveis efeitos de interação entre os alelos do fundador k e do fundador k' e tem dimensão 3 se $k = k'$ e dimensão 4 se $k \neq k'$. De maneira similar à variância de (a_k^j) , a variância de $d_{kk'}^j$ depende do coeficiente de endogamia dos fundadores k e k' e se reduz a $\sigma_{d_j}^2 I$ se estes coeficientes forem considerados iguais a zero.

As distribuições a priori atribuídas por Yi & Xu (2001) aos componentes de variância, $\sigma_{a_j}^2$, $\sigma_{d_j}^2$ e σ_e^2 , foram todas uniformes, embora outras distribuições, em especial da família Gama, possam ser utilizadas.

As matrizes Z^m e Z^p apresentam distribuição condicional dada por $p(Z^m, Z^p | l, \lambda, M)$, ou seja, dados o número de QTL, as suas posições e a informação dos marcadores, tem-se a probabilidade de ocorrer o genótipo do QTL, especificado em tais matrizes.

2.6.3 Distribuições a posteriori

A distribuição conjunta a posteriori de todos os parâmetros, dado o vetor y , $p(\theta|y)$, é especificada pelo produto de todas as prioris e a função

de verossimilhança.

$$\begin{aligned}
p(\theta|y) &\propto p(y|\theta)p(Z^m, Z^p|l, \lambda, M)p(l)p(b)p(\sigma_e^2) \times \\
&\times \prod_{j=1}^l (p(a_j|\sigma_{a_j}^2)p(\sigma_{a_j}^2)p(d_j|\sigma_{d_j}^2)p(\sigma_{d_j}^2)p(\lambda_j)). \quad (11)
\end{aligned}$$

A partir desta expressão, após terem sido substituídas todas as priors e a verossimilhança, são obtidas as distribuições condicionais completas a posteriori de cada parâmetro. Com base nessas distribuições, os valores dos parâmetros podem ser atualizados a cada iteração, por meio dos métodos MCMC apropriados em cada caso.

As atualizações dos parâmetros, segundo Yi & Xu (2001), podem ser feitas utilizando-se o algoritmo do amostrador de Gibbs para a matriz do genótipo completo das marcas, considerada parcialmente desconhecida em seu trabalho, e para as matrizes de herança alélica Z^m e Z^p . Para os demais parâmetros, foi utilizado o algoritmo Metropolis-Hastings, incluindo um passo do algoritmo MCMC com saltos reversíveis para a atualização do número de QTLs.

2.6.4 Probabilidades de aceitação

Os valores candidatos em cada passo do algoritmo Metropolis-Hastings são aceitos com base em probabilidades de aceitação que dependem das distribuições condicionais completas e das distribuições geradoras de candidatos especificadas para cada parâmetro.

No processo de atualização apresentado por Yi & Xu (2001), as posições do j -ésimo QTL são atualizadas, uma por vez, e, juntamente com cada posição λ_j , são atualizadas no mesmo passo as matrizes de herança alélica, de dominância e dos indicadores de segregação (U_j^m, U_j^p) . Os valores candidatos λ_j^* , (U_j^{m*}, U_j^{p*}) , Z_j^{m*} , Z_j^{p*} e W_j^* são aceitos com probabilidade dada pelo mínimo entre 1 e A , sendo

$$A = \frac{p(y|\theta^*, \lambda_j^*, Z_j^{m*}, Z_j^{p*}) p(U_j^{m*}, U_j^{p*} | \lambda_j^*, G_j^E, G_j^D)}{p(y|\theta)} \frac{q(U_j^m, U_j^p)}{p(U_j^m, U_j^p | \lambda_j, G_j^E, G_j^D) q(U_j^{m*}, U_j^{p*})}. \quad (12)$$

Em (12), G_j^E, G_j^D denotam o genótipo marcador de todos os membros do pedigree à esquerda e à direita da posição λ_j ; $q(U_j^m, U_j^p)$ denota a distribuição geradora de candidatos dos indicadores de segregação; θ^* contém todos os elementos de θ , exceto λ_j, Z_j^m, Z_j^p e $p(U_j^{m*}, U_j^{p*} | \lambda_j^*, G_j^E, G_j^D)$ é a distribuição a posteriori condicional para o genótipo (U_j) do j-ésimo QTL.

A atualização dos efeitos alélicos do j-ésimo QTL é feita considerando-se um fundador por vez, gerando um valor candidato para o vetor a_k^j e aceitando ou não esse novo valor com base na correspondente probabilidade de aceitação dada por mínimo entre 1 e A , sendo,

$$A = \frac{p(y|\theta^*)p(a_k^{j*} | \sigma_{aj}^2)}{p(y|\theta)p(a_k^j | \sigma_{aj}^2)}. \quad (13)$$

Em (13), θ^* contém todos os elementos de θ , com o vetor a_k^j substituído por a_k^{j*} , e $p(a_k^j | \sigma_{aj}^2)$ é a distribuição a priori dos efeitos alélicos do k-ésimo fundador.

Os efeitos de dominância do QTL são atualizados para dois fundadores por vez, ou seja, um valor candidato do vetor $d_{kk'}^j$, que representa os efeitos de dominância obtidos pela interação entre os alelos dos fundadores k e k' , é aceito com probabilidade dada pelo mínimo entre 1 e A , sendo,

$$A = \frac{p(y|\theta^*)p(d_{kk'}^{j*} | \sigma_{dj}^2)}{p(y|\theta)p(d_{kk'}^j | \sigma_{dj}^2)}. \quad (14)$$

Em (14), θ^* contém todos os elementos de θ , com o vetor $d_{kk'}^j$ substituído por $d_{kk'}^{j*}$, $p(d_{kk'}^j | \sigma_{dj}^2)$ é a distribuição a priori para os efeitos de do-

minância em $d_{kk'}^j$.

O valor candidato para a variância alélica do j-ésimo QTL $\sigma_{a_j}^{2*}$ é simulado de uma uniforme em torno do valor atual $\sigma_{a_j}^2$ e aceito com probabilidade dada pelo mínimo entre 1 e A , sendo

$$A = \frac{\prod_{k=1}^m p(a_k^j | \sigma_{a_j}^{2*})}{\prod_{k=1}^m p(a_k^j | \sigma_{a_j}^2)}. \quad (15)$$

A variância de dominância do j-ésimo QTL é atualizada pelo mesmo processo que a variância aditiva, sendo o valor candidato $\sigma_{d_j}^{2*}$ aceito com probabilidade dada pelo mínimo entre 1 e A , em que

$$A = \frac{\prod_{k=1}^m \prod_{k'=k}^m p(d_{kk'}^j | \sigma_{d_j}^{2*})}{\prod_{k=1}^m \prod_{k'=k}^m p(d_{kk'}^j | \sigma_{d_j}^2)}. \quad (16)$$

A atualização do parâmetro l é baseada no algoritmo Metropolis Hastings com saltos reversíveis, no qual a decisão entre tentar adicionar um novo QTL ao modelo ou tentar excluir um QTL já existente é tomada com base em uma probabilidade definida previamente a cada iteração. A probabilidade p_a de tentar adicionar um novo QTL ao modelo é definida como sendo $p_a = 0$ se $l = l_{maximo}$ e $p_a = 1$ se $l = 1$. Para os demais casos, $p_a = 0,5$. A probabilidade de se tentar excluir um QTL existente no modelo é definida como $p_e = 1 - p_a$.

Se a decisão de tentar adicionar um novo QTL ao modelo com l QTLs é tomada, uma nova posição, λ_{l+1} , é simulada, e todos os demais parâmetros, a_{l+1} , d_{l+1} , $\sigma_{a_{l+1}}^2$, $\sigma_{d_{l+1}}^2$, Z_{l+1}^m , Z_{l+1}^p , referentes a esse novo QTL são gerados por meio das respectivas distribuições candidatas especificadas nos processos de atualização. A adição do novo QTL é aceita com probabilidade dada pelo mínimo entre 1 e A (Yi & Xu, 2001), sendo

$$A = \frac{p_e}{p_a} \times \frac{p(y|\theta^*)}{p(y|\theta)} \frac{p(U_{l+1}^m, U_{l+1}^p | \lambda_{l+1}, G_{l+1}^E, G_{l+1}^D)}{(l+1)^2} \frac{l_0}{q(U_{l+1}^m, U_{l+1}^p)}, \quad (17)$$

em que $\theta^* = (\theta, \lambda_{l+1}, a_{l+1}, d_{l+1}, \sigma_{a_{l+1}}^2, \sigma_{d_{l+1}}^2, Z_{l+1}^m, Z_{l+1}^p)$ com l em θ substituído por $l+1$, G_{l+1}^E e G_{l+1}^D denotam o genótipo marcador à esquerda e à direita da posição λ_{l+1} , para todos os membros do pedigree. Se a adição do novo QTL é aceita, o novo modelo fica com $l+1$ QTLs. Caso contrário, o modelo permanece com l QTLs.

Se a decisão de tentar excluir um dos l QTLs já existentes no modelo é tomada, a escolha entre um deles é feita aleatoriamente. Se o j -ésimo QTL é escolhido, então, ele é excluído do modelo, com probabilidade dada pelo mínimo entre 1 e A (Yi & Xu, 2001).

$$A = \frac{p_a}{p_e} \times \frac{p(y|\theta^*)}{p(y|\theta)} \times \frac{l^2}{p(U_j^m, U_j^p | \lambda_j, G_j^E, G_j^D)} \frac{q(U_j^m, U_j^p)}{l_0}. \quad (18)$$

Em (18), θ^* contém todos os elementos de θ , exceto aqueles correspondentes ao j -ésimo QTL. Se a exclusão do QTL é aceita, o modelo fica com $l-1$ QTLs. Caso contrário, o modelo permanece com l QTLs.

As expressões das probabilidades de aceitação para adicionar ou remover um QTL do modelo foram apresentadas em várias versões, de acordo com as suposições de cada estudo (Heath, 1997; Uimari & Hoeschele, 1997; Sillanpää & Arjas, 1998; Stephens & Fisch, 1998). De acordo com Jannink & Fernando (2004), se o vetor de parâmetros não é ordenado, algumas das suposições feitas na derivação da expressão final dessas probabilidades não são válidas, ocasionando riscos de erro nos resultados. Uma ampla discussão dos possíveis erros de cada expressão e da derivação correta das probabilidades é apresentada por esses autores, mostrando que os termos referentes aos números de QTLs nos modelos atual e candidato não devem aparecer

elevados ao quadrado, como ocorre nas expressões (17) e (18) e também na maioria das expressões apresentadas em estudos anteriores.

3 METODOLOGIA

3.1 Simulações

Foi construído um programa de simulação de pedigrees complexos completamente geral, permitindo a suposição de diferentes números de fundadores (nf), números de filhos por cruzamento (nfc), números de QTLs (nQTL) e números de cromossomos (nc), mapeados a distâncias conhecidas (dM). São sorteadas duas gerações com nf cruzamentos cada, permitindo autofecundações.

Modelos para pedigrees com elevado número de indivíduos, como considerado por Yi & Xu (2001), demandam grande esforço computacional. Para superar essa dificuldade, foram construídas as três simulações a seguir.

3.1.1 Simulação 1 (nf=5,nfc=6,nc=2,nQTL=3,dM=10)

A primeira simulação teve por objetivo verificar a viabilidade do algoritmo RJ. O pedigree complexo originou-se de 5 cruzamentos aleatórios entre 5 indivíduos fundadores. Devido à aleatorização, alguns indivíduos são mais bem representados na próxima geração e outros nem são representados. De cada um dos 5 cruzamentos, foram gerados 6 indivíduos, totalizando, assim, 30 indivíduos na primeira geração. Desta geração, 5 indivíduos foram selecionados aleatoriamente para serem os pais e 5 para serem as mães e, novamente, foram feitos 5 cruzamentos aleatórios entre eles, originando, assim, os 30 indivíduos da segunda geração .

Para os 65 indivíduos do pedigree foram simulados valores de uma característica quantitativa controlada por 3 QTLs com múltiplos alelos por loco, sendo 2 localizados no primeiro cromossomo, de comprimento 100cM, e 1 localizado no segundo cromossomo, de comprimento 70cM. Nos dois cromossomos, foram simulados genes marcadores a cada 10cM, com 6 alelos igualmente prováveis para cada loco marcador, totalizando 11 marcadores no cromossomo 1, e 8 marcadores no cromossomo 2. O pedigree resultante

desta simulação pode ser verificado na Figura 1. Na primeira linha, os círculos representam os fundadores; na segunda, os quadrados numerados representam cada indivíduo da primeira geração, sendo que cada bloco representa uma família. A segunda linha de círculos representa os indivíduos da primeira geração que são os pais da segunda geração. Na última linha, os quadrados representam os indivíduos da segunda geração de filhos.

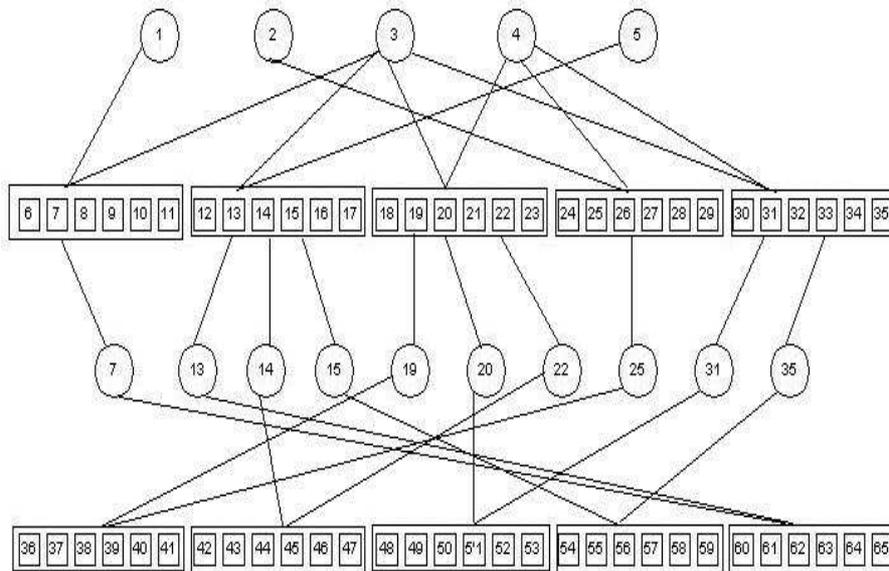


FIGURA 1: Pedigree simulado, consistindo de 65 indivíduos, sendo uma geração de fundadores e duas gerações de descendentes.

Os 10 efeitos alélicos, referentes aos alelos de todos os fundadores, e os 55 efeitos de dominância, referentes à interação entre pares de alelos dos fundadores, de cada QTL foram gerados de distribuições normais com média zero e variâncias aditiva (alélica) e de dominância dadas na Tabela 1 a seguir.

O efeito fixo b (média geral) foi simulado com valor zero. Os erros foram simulados de uma distribuição normal com média zero e variância 1 ($\sigma_e^2 = 1$).

TABELA 1: Variâncias alélicas e de dominância simuladas para os 3 QTLs com as respectivas posições no cromossomo.

| Cromossomo | posição (cM) | v. alélica | v. de dominância |
|------------|--------------|------------|------------------|
| 1 | 25 | 0,15 | 0,10 |
| 1 | 75 | 0,10 | 0,05 |
| 2 | 25 | 0,15 | 0,10 |

A análise desse conjunto de dados baseou-se em 100.000 amostras válidas das cadeias de cada parâmetro. Esta amostra foi obtida combinando-se duas cadeias em paralelo de 48.000 iterações cada, com descarte inicial de 2.000 valores e salto de 10, com uma terceira cadeia com 4.000 amostras válidas, obtida após um descarte de 1.000 iterações iniciais e salto de 25.

Em uma segunda análise desta simulação, considerou-se o número de QTLs como sendo conhecido, realizando-se, portanto, inferências condicionais a esse número. As cadeias com 50.000 iterações foram obtidas após um descarte de 1.000 iterações iniciais e um salto entre amostras de 10 iterações. Esta nova análise teve a finalidade de verificar a eficiência da estimação das posições de QTLs, com o método considerado.

3.1.2 Simulação 2 (nf=5,nfc=10,nc=1,nQTL=2,dM=10)

Esta simulação teve por objetivo verificar a influência da redução do número de cromossomos para apenas um, e do pequeno aumento no número de indivíduos amostrados, nas estimativas das posições dos QTLs, e nas propriedades de mistura e estimação do algoritmo RJ. Em relação à primeira simulação, aumentou-se para 10 o número de filhos por cruzamento, obtendo-se, de forma similar, um total de 105 indivíduos no pedigree. Os valores fenotípicos foram simulados sob a influência de 2 QTLs multi-alélicos localizados nas posições 34 e 75cM de um único cromossomo de comprimento 100cM. A cada 10cM do cromossomo, foram simulados genes marcadores, sendo que para cada indivíduo foram sorteados 2 de 6 alelos equiprováveis por loco.

Os efeitos aditivos e de dominância dos dois QTLs foram simulados de distribuições normais com média zero e variâncias dadas por 0,15 e 0,10 para as variâncias aditivas, e para as variâncias de dominância 0,10 e 0,05. Os valores para a média geral e variância residual foram os mesmos da simulação anterior.

As distribuições a posteriori da análise desse conjunto de dados foram obtidas descartando-se as 1.000 iterações iniciais e tomando-se, em seguida, amostras a cada 10 iterações, até obter-se cadeias válidas de tamanho 9.000.

3.1.3 Simulação 3 (nf=5,nfc=10,nc=1,nQTL=2,dM=2)

Em uma terceira simulação, optou-se por aumentar o número de marcadores por cromossomo, como forma de obter mais informação a respeito da localização e efeitos dos QTLs, já que um aumento no tamanho do pedigree, que seria desejável, dificulta a análise devido ao tempo computacional exigido.

O experimento manteve o mesmo delineamento da simulação 2, com 105 indivíduos, considerando da mesma forma um único cromossomo de comprimento 100cM, com dois QTLs simulados nas posições 33 e 71cM. As variâncias aditiva e de dominância desses dois QTLs foram simuladas com os mesmos valores da simulação anterior, assim como a variância residual e a média geral. Os locos marcadores foram simulados da mesma forma, porém, a cada 2cM, obtendo-se, assim, um total de 51 marcadores no cromossomo.

As distribuições a posteriori da análise desse conjunto de dados foram obtidas descartando-se as 1.000 iterações iniciais e tomando-se, em seguida, amostras a cada 10 iterações, até obter-se cadeias com 24.000 amostras válidas.

As simulações e as análises de todas as situações consideradas foram realizadas por meio de rotinas elaboradas para o software R (R Development Core Team, 2008), apresentadas no anexo A.

3.2 Modelo linear misto e análise bayesiana

Os dados simulados, supondo uma característica controlada por l QTLs, foram descritos de acordo com o seguinte modelo linear:

$$y = Xb + \sum_{j=1}^l (Z_j^m + Z_j^p)a_j + \sum_{j=1}^l W_j d_j + e. \quad (19)$$

Os termos do modelo são definidos como na seção 2.6.1.

As matrizes de herança alélica Z_j^m e Z_j^p são construídas, conforme apresentado por Yi & Xu (2001), considerando para o j -ésimo QTL a sua posição λ_j , e os indicadores de segregação materno e paterno do QTL, u_{ij}^m e u_{ij}^p , que são sorteados para todos os indivíduos das gerações filiais. Os dois alelos de cada indivíduo são identificados como alelo materno e alelo paterno. Os possíveis valores dos indicadores de segregação são $(u_{ij}^m, u_{ij}^p) = (1,1), (1,0), (0,1)$ ou $(0,0)$, sendo que $u_{ij}^m = 1$ indica que o indivíduo i recebeu o alelo materno da mãe, com relação ao j -ésimo QTL e $u_{ij}^m = 0$ indica que o indivíduo i não recebeu o alelo materno da mãe para o j -ésimo QTL, ou seja, ele recebeu o alelo paterno da mãe. Similarmente, o indicador de segregação paterno $u_{ij}^p = 1$ indica que o indivíduo i recebeu o alelo materno do pai e $u_{ij}^p = 0$ indica que o indivíduo i não recebeu o alelo materno do pai. Cada vetor dos indicadores de segregação (u_{ij}^m, u_{ij}^p) compõe uma linha das matrizes (U_j^m, U_j^p) .

Com base nas matrizes dos indicadores de segregação, (U_j^m, U_j^p) , são construídas as matrizes de herança alélica, supondo que os indivíduos no pedigree estejam ordenados de forma que os progenitores estejam sempre antes de suas progênes. Se os $2m$ alelos dos fundadores são numerados de 1 a $2m$, então, os alelos $2k - 1$ e $2k$ são os alelos do k -ésimo fundador. Os vetores $Z_j^m(i)$ e $Z_j^p(i)$ denotam a i -ésima linha de Z_j^m e Z_j^p , respectivamente, ou seja, referem-se às identificações dos alelos materno e paterno do i -ésimo indivíduo, para o j -ésimo QTL.

Supondo que as primeiras $(i - 1)$ linhas de Z_j^m e Z_j^p já foram cons-

truídas, a construção da i -ésima linha segue os passos a seguir. Se o i -ésimo indivíduo é um dos m fundadores (por exemplo, o k -ésimo fundador), então, a posição $2k - 1$ de $Z_j^m(i)$ e a posição $2k$ de $Z_j^p(i)$ recebem o valor 1, e as demais posições dessas linhas recebem o valor zero. Mas, se o i -ésimo indivíduo não é um fundador, e sim progênie dos indivíduos i_1 (mãe) e i_2 (pai), a i -ésima linha de Z_j^m e Z_j^p é construída conforme as expressões a seguir.

$$\begin{aligned} Z_j^m(i) &= u_{ij}^m Z_j^m(i_1) + (1 - u_{ij}^m) Z_j^p(i_1) ; \\ Z_j^p(i) &= u_{ij}^p Z_j^m(i_2) + (1 - u_{ij}^p) Z_j^p(i_2) \end{aligned}$$

Esse processo é repetido para todas as linhas até que as matrizes estejam completas e, em seguida, a partir dessas matrizes, a matriz W_j também é obtida. Cada linha de W_j apresenta o valor 1 apenas na posição correspondente ao efeito de dominância obtido pela interação entre os dois alelos do indivíduo correspondente, sendo todos os demais elementos dessa linha iguais a zero.

A função de verossimilhança, supondo a característica normalmente distribuída, é dada pela expressão (10).

A matriz de marcadores M foi assumida conhecida para todas as observações, ou seja, para todos os indivíduos tem-se a informação da posição e do genótipo de todos os locos marcadores.

As distribuições a priori para cada parâmetro foram baseadas naquelas já consideradas em trabalhos anteriores. Para o número de QTLs l adotou-se uma distribuição a priori Poisson truncada, com valor máximo ($l_{maximo} = 6$) e média $l_0 = 2$. As distribuições a priori assumidas para a posição do QTL (λ_j) e os efeitos fixos (b) foram uniformes. Os efeitos aleatórios a_j e d_j foram considerados a priori com distribuições normais, cujas expressões foram especificadas no item (2.6.2). Para os componentes de variância ($\sigma_{a_j}^2$) e ($\sigma_{d_j}^2$), assumiram-se distribuições a priori qui-quadrado invertidas escaladas com 25 graus de liberdade e parâmetro de escala igual a 3 e, para a variância ambiental (σ_e^2), qui-quadrado invertida com 8 graus

de liberdade e parâmetro de escala igual a 10.

3.2.1 Distribuições a posteriori

A distribuição conjunta a posteriori $p(\theta|y)$ é obtida combinando-se a informação a priori de todos os parâmetros e a função de verossimilhança.

$$\begin{aligned}
p(\theta|y) &\propto (\sigma_e^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma_e^2}R^\top R\right\}p(Z^m, Z^p|l, \lambda, M) \\
&\quad \prod_{j=1}^l \left(\prod_{k=1}^m |var(a_k^j)|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(a_k^j)^\top (var(a_k^j))^{-1}(a_k^j)\right\} \right. \\
&\quad \left. \prod_{k=1}^m \prod_{k'=k}^m |var(d_{kk'}^j)|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(d_{kk'}^j)^\top (var(d_{kk'}^j))^{-1}(d_{kk'}^j)\right\} \right) \\
&\quad p(\sigma_{d_j}^2)p(\sigma_{a_j}^2)p(\lambda_j) \Big) \frac{l_0^l}{l!} \exp\{-l_0\}p(b)p(\sigma_e^2) \quad (20)
\end{aligned}$$

As distribuições condicionais completas a posteriori de cada um dos parâmetros foram obtidas a partir da expressão da distribuição conjunta e estão apresentadas a seguir.

A distribuição condicional completa para o parâmetro λ foi obtida individualmente para cada posição λ_j do j -ésimo QTL e é especificada pela seguinte expressão.

$$\begin{aligned}
p(\lambda_j|\theta^*, y) &\propto \exp\left\{-\frac{1}{2\sigma_e^2}R^\top R\right\}p(Z_j^m, Z_j^p|l, \lambda_j, M)p(\lambda_j) \\
&\propto \exp\left\{-\frac{1}{2\sigma_e^2}R^\top R\right\}p(U_j^m, U_j^p|\lambda_j, G_j^E, G_j^D)p(\lambda_j) \quad (21)
\end{aligned}$$

Em (21), G_j^E e G_j^D denotam o genótipo marcador de todos os indivíduos à esquerda e à direita, respectivamente, dos locos flanqueadores da posição λ_j do j -ésimo QTL; θ^* denota o vetor θ com todos os elementos

exceto a posição λ_j ; R é definido como na expressão (10); U_j^m e U_j^p são os vetores (de ordem $n - m$) dos indicadores de segregação materno e paterno para o j -ésimo QTL na correspondente posição; $p(U_j^m, U_j^p | \lambda_j, G_j^E, G_j^D)$ é a distribuição a posteriori condicional para o genótipo (U_j) do j -ésimo QTL.

A probabilidade a posteriori do genótipo U_j é calculada para o i -ésimo indivíduo, de acordo com a expressão a seguir (Xu & Yi, 2000):

$$p(U_{ij} = k | y_i, \lambda_j, G_{ij}^E = t, G_{ij}^D = s) = \frac{p(y_i | \theta) p(G_{ij}^E = t, G_{ij}^D = s | \lambda_j, U_{ij} = k) p(U_{ij} = k)}{\sum_{k=1}^4 p(y_i | \theta) p(G_{ij}^E = t, G_{ij}^D = s | \lambda_j, U_{ij} = k) p(U_{ij} = k)}, \quad (22)$$

sendo $k, t, s = 1, \dots, 4$, com k representando os 4 possíveis genótipos do QTL (indicadores de segregação) e t, s representando os 4 possíveis genótipos dos marcadores à esquerda e à direita, respectivamente, da posição λ_j . Condicionalmente sobre U_{ij} , G_{ij}^E e G_{ij}^D são independentes e, assim, $p(G_{ij}^E = t, G_{ij}^D = s | \lambda_j, U_{ij} = k) = p(G_{ij}^E = t | \lambda_j, U_{ij} = k) \times p(G_{ij}^D = s | \lambda_j, U_{ij} = k)$. A probabilidade condicional do genótipo marcador à esquerda ($p(G_{ij}^E = t | \lambda_j, U_{ij} = k)$) ou à direita ($p(G_{ij}^D = s | \lambda_j, U_{ij} = k)$) é obtida a partir da matriz de transição P_{GU} a seguir, na qual r é a fração de recombinação entre o QTL e o marcador, à esquerda ou à direita.

$$P_{GU} = \begin{bmatrix} (1-r)^2 & (1-r)r & (1-r)r & r^2 \\ (1-r)r & (1-r)^2 & r^2 & (1-r)r \\ (1-r)r & r^2 & (1-r)^2 & (1-r)r \\ r^2 & (1-r)r & (1-r)r & (1-r)^2 \end{bmatrix}$$

Os efeitos aditivos do j -ésimo QTL possuem distribuição condicional completa para cada vetor a_k^j dada por:

$$p(a_k^j | \theta^*, y) \propto \exp\left\{-\frac{1}{2\sigma_e^2} R^\top R\right\} \exp\left\{-\frac{1}{2}(a_k^j)^\top (\text{var}(a_k^j))^{-1} (a_k^j)\right\}, \quad (23)$$

em que θ^* é o vetor com todos os elementos de θ , com exceção dos efeitos aditivos referentes ao k -ésimo fundador e R é definido como na expressão (10).

A distribuição condicional completa para os efeitos de dominância é obtida para o vetor $d_{kk'}^j$ em cada um dos l QTLs para os fundadores k e k' .

$$p(d_{kk'}^j | \theta^*, y) \propto \exp\left\{-\frac{1}{2\sigma_e^2} R^\top R\right\} \times \exp\left\{-\frac{1}{2}(d_{kk'}^j)^\top (\text{var}(d_{kk'}^j))^{-1} (d_{kk'}^j)\right\}, \quad (24)$$

sendo θ^* o vetor com todos os elementos de θ , com exceção dos efeitos de dominância referentes aos fundadores k e k' e R é definido como na expressão (10).

A variância aditiva do j -ésimo QTL, $\sigma_{a_j}^2$, tem distribuição condicional completa dada por:

$$\begin{aligned} p(\sigma_{a_j}^2 | \theta^*, y) &\propto p(\sigma_{a_j}^2) \prod_{k=1}^m |\text{var}(a_k^j)|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(a_k^j)^\top (\text{var}(a_k^j))^{-1} (a_k^j)\right\} \\ &\propto (\sigma_{a_j}^2)^{-\frac{2m}{2}-1} \exp\left\{-\frac{1}{2\sigma_{a_j}^2} (a_j)^\top (a_j)\right\} \\ &\times (\sigma_{a_j}^2)^{-\frac{h_a}{2}-1} \exp\left\{-\frac{S_a}{2\sigma_{a_j}^2}\right\} \\ &\propto (\sigma_{a_j}^2)^{-\frac{(h_a+2m)}{2}-1} \exp\left\{-\frac{1}{2\sigma_{a_j}^2} ((a_j)^\top a_j + S_a)\right\} \end{aligned} \quad (25)$$

A distribuição condicional completa para a variância aditiva tem a forma de uma distribuição qui-quadrado escalada invertida com $(h_a + 2m)$ graus de liberdade e parâmetro de escala $((a_j)^\top a_j + S_a)$.

A variância de dominância do j -ésimo QTL $\sigma_{d_j}^2$ tem uma distribuição condicional completa dada pela expressão apresentada a seguir.

$$\begin{aligned}
p(\sigma_{d_j}^2 | \theta^*, y) &\propto p(\sigma_{d_j}^2) \prod_{k=1}^m \prod_{k'=k}^m |var(d_{kk'}^j)|^{-\frac{1}{2}} \\
&\times \exp\left\{ -\frac{1}{2} (d_{kk'}^j)^\top (var(d_{kk'}^j))^{-1} (d_{kk'}^j) \right\} \\
&\propto (\sigma_{d_j}^2)^{-\frac{(ha+nd)}{2}-1} \exp\left\{ -\frac{((d_j)^\top d_j + S_d)}{2\sigma_{d_j}^2} \right\} \quad (26)
\end{aligned}$$

A distribuição condicional completa para a variância de dominância tem a forma de uma distribuição qui-quadrado escalada invertida com $(ha + nd)$ graus de liberdade e parâmetro de escala $(d_j)^\top d_j + S_d$.

Nas expressões (25) e (26), as distribuições a priori $p(\sigma_{a_j}^2)$ e $p(\sigma_{d_j}^2)$ foram consideradas qui-quadrado invertidas escaladas, m é o número de fundadores, sendo $nd = m(2m + 1)$, e o vetor θ^* indica o vetor de todos os parâmetros, com exceção da variância aditiva e de dominância, respectivamente.

Os efeitos fixos b têm distribuição condicional completa dada pela expressão a seguir.

$$p(b | \theta^*, y) \propto p(b) \exp\left\{ -\frac{1}{2\sigma_e^2} R^\top R \right\}, \quad (27)$$

sendo $R = y - Xb - \sum_{j=1}^l (Z_j^m + Z_j^p) a_j - \sum_{j=1}^l W_j d_j$.

A partir da expressão (27), lembrando que $p(b)$ é constante, verifica-se que Xb tem distribuição normal com média $\left(y - \sum_{j=1}^l (Z_j^m + Z_j^p) a_j - \sum_{j=1}^l W_j d_j \right)$ e variância σ_e^2 , sendo θ^* o vetor de parâmetros com exceção de b . Desta última expressão, após algumas manipulações algébricas, conclui-se que b tem distribuição normal com média dada por $(X^\top X)^{-1} X^\top \left(y - \sum_{j=1}^l (Z_j^m + Z_j^p) a_j - \sum_{j=1}^l W_j d_j \right)$ e variância dada por $(X^\top X)^{-1} \sigma_e^2$.

Para a variância ambiental σ_e^2 , a distribuição condicional completa

a posteriori obtida tem a forma apresentada a seguir.

$$\begin{aligned}
p(\sigma_e^2 | \theta^*, y) &\propto (\sigma_e^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma_e^2} R^\top R\right\} p(\sigma_e^2) \\
&\propto \left(\frac{1}{\sigma_e^2}\right)^{\frac{n+he}{2}+1} \exp\left\{-\frac{1}{2\sigma_e^2}(R^\top R + Se)\right\} \quad (28)
\end{aligned}$$

Verifica-se, por meio da expressão (28), que a condicional completa para a variância ambiental tem a forma de uma distribuição qui-quadrado escalada invertida com $(n + he)$ graus de liberdade e parâmetro de escala $(R^\top R + Se)$, sendo R definido como anteriormente.

As distribuições condicionais completas apresentadas não possuem formas de densidades conhecidas apenas para os parâmetros λ , a e d , como pode ser visto nas expressões (21) a (24). Para os demais parâmetros, as distribuições condicionais completas (expressões de 25 a 28) apresentam formas de densidades conhecidas.

As amostras das distribuições condicionais completas de cada parâmetro foram obtidas por meio de algoritmos (MCMC).

3.3 Algoritmos de amostragem

O algoritmo do amostrador de Gibbs foi usado para amostrar as distribuições condicionais completas dos parâmetros média geral, b e variância residual σ_e^2 , e para as componentes de variância dos QTLs, por estas serem densidades conhecidas e fáceis de serem amostradas.

Para os demais parâmetros, a amostragem das distribuições condicionais completas foi feita utilizando-se o algoritmo Metropolis Hastings, considerando, para cada caso, uma distribuição candidata apropriada para o processo de atualização a cada iteração, com as respectivas probabilidades de aceitação para os valores candidatos.

As posições de cada QTL são atualizadas uma por vez e, juntamente

com cada posição λ_j , são atualizadas no mesmo passo as matrizes $U_j^m, U_j^p, Z_j^m, Z_j^p$ e W_j (Yi & Xu, 2001). De acordo com o método de amostragem considerado, a ordem dos QTLs não é fixada. Para o j -ésimo QTL, um valor candidato λ_j é amostrado de uma distribuição uniforme em torno do valor atual, ou seja, do intervalo $(\lambda_j - \alpha, \lambda_j + \alpha)$. Ao mesmo tempo, novos valores dos indicadores de segregação (U_j^m, U_j^p) são gerados diretamente de uma distribuição discreta, podendo esses indicadores, (u_{ij}^m, u_{ij}^p) , para cada indivíduo i , assumir um dos quatro valores $(1,1), (1,0), (0,1)$ ou $(0,0)$. Com base nesses indicadores de segregação, são geradas matrizes candidatas, Z_j^{m*} e Z_j^{p*} e, posteriormente, uma nova matriz candidata W_j^* também é obtida. Os valores candidatos são aceitos com probabilidade dada de acordo com a expressão (12). No intervalo da uniforme candidata para λ_j , definiu-se o valor de α igual a 0,15M em todas as análises realizadas. As inferências sobre as posições dos QTLs foram baseadas no método da intensidade a posteriori de QTL, definida na expressão (5).

A atualização dos efeitos alélicos dos QTL é feita gerando um valor candidato para o vetor a_k^j do j -ésimo QTL e aceitando ou não esse novo valor com base na correspondente probabilidade de aceitação. Para tanto, duas variáveis aleatórias δ_1 e δ_2 , são simuladas de uma distribuição uniforme em torno de zero, com comprimento fixo predeterminado, de forma que resulte em uma taxa razoável de aceitação. Nas análises das duas estratégias de simulação, utilizou-se o intervalo $(-0,8, 0,8)$. O valor candidato a_k^{j*} é obtido, então, por $a_k^{j*} = a_k^j + (\delta_1, f_k \delta_1 + (1 - f_k) \delta_2)^\top$, sendo f_k o coeficiente de endogamia do k -ésimo fundador (Yi & Xu, 2001). A probabilidade de aceitação desse valor candidato é obtida da expressão (13).

Os efeitos de dominância do QTL são atualizados para dois fundadores por vez, ou seja, um valor candidato é gerado para o vetor $d_{kk'}^j$, que representa os efeitos de dominância obtidos pela interação dos alelos dos fundadores k e k' e a probabilidade de aceitação é verificada para cada um desses vetores do j -ésimo QTL. De acordo com o método proposto por Yi &

Xu (2001), variáveis aleatórias $\delta_1, \delta_2, \delta_3$ e δ_4 são simuladas de uma uniforme simétrica em torno de zero. O mesmo intervalo da amostragem dos efeitos aditivos foi utilizado nessa atualização. O vetor candidato $d_{kk'}^{j*}$, quando $k \neq k'$, é então obtido por:

$$d_{kk'}^{j*} = d_{kk'}^j + \left(\delta_1, f_k \delta_1 + (1 - f_k) \delta_2, f_{k'} \delta_1 + (1 - f_{k'}) \delta_3, f_k f_{k'} \delta_1 + (1 - f_k) f_{k'} \delta_2 + f_k (1 - f_{k'}) \delta_3 + (1 - f_k) (1 - f_{k'}) \delta_4 \right)^\top.$$

Para o caso em que $k = k'$, tem-se:

$$d_{kk'}^{j*} = d_{kk'}^j + \left(\delta_1, f_k \delta_1 + (1 - f_k) \delta_2, f_k \delta_1 + (1 - f_k) \delta_4 \right)^\top.$$

O valor candidato é aceito com probabilidade obtida a partir da expressão (14).

Nos exemplos de simulação utilizados, os fundadores foram considerados não endogâmicos, sendo f_k igual a zero.

A atualização das componentes de variância é feita amostrando-se cada componente diretamente de sua distribuição condicional completa a posteriori, utilizando-se o algoritmo do amostrador de Gibbs.

A atualização do parâmetro l é baseada no algoritmo Metropolis Hastings com saltos reversíveis, no qual a decisão entre tentar adicionar um novo QTL ao modelo ou tentar excluir um QTL já existente é tomada com base em uma probabilidade definida previamente a cada iteração. A probabilidade p_a de adicionar um novo QTL ao modelo é definida como sendo $p_a = 0$ se $l = l_{maximo}$ e $p_a = 1$ se $l = 1$, não sendo, portanto, permitidos modelos com nenhum QTL. Para os demais casos, $p_a = 0,5$. A probabilidade de excluir um QTL existente no modelo é definida como $p_e = 1 - p_a$.

As probabilidades de aceitação para incluir ou excluir um QTL do modelo foram definidas com base naquelas apresentadas por Yi & Xu (2001), acrescentando as mudanças propostas por Jannink & Fernando (2004). Se é tomada a decisão de tentar adicionar um novo QTL ao modelo atual com

l QTLs, uma nova posição, λ_{l+1} , é simulada, e todos os demais parâmetros, a_{l+1} , d_{l+1} , $\sigma_{a_{l+1}}^2$, $\sigma_{d_{l+1}}^2$, Z_{l+1}^m , Z_{l+1}^p , referentes a esse novo QTL, são gerados por meio das respectivas distribuições candidatas especificadas nos processos de atualização. A adição do novo QTL é aceita com probabilidade dada pelo mínimo entre 1 e A (Yi & Xu, 2001; Jannink & Fernando, 2004), sendo

$$A = \frac{p_e}{p_a} \times \frac{p(y|\theta^*)}{p(y|\theta)} \frac{p(U_{l+1}^m, U_{l+1}^p | \lambda_{l+1}, G_{l+1}^E, G_{l+1}^D)}{l+1} \frac{l_0}{q(U_{l+1}^m, U_{l+1}^p)}, \quad (29)$$

em que $\theta^* = (\theta, \lambda_{l+1}, a_{l+1}, d_{l+1}, \sigma_{a_{l+1}}^2, \sigma_{d_{l+1}}^2, Z_{l+1}^m, Z_{l+1}^p)$ com l em θ substituído por $l+1$, G_{l+1}^E e G_{l+1}^D denotam o genótipo marcador completo de todos os membros do pedigree, à esquerda e à direita da posição λ_{l+1} . Se for aceita a adição do novo QTL, aceita-se, então, o novo modelo com $(l+1)$ QTLs e vetor de parâmetros θ^* . Caso contrário, o modelo permanece com l QTLs e vetor de parâmetros θ .

Se a decisão de tentar excluir um QTL do modelo com l QTLs é tomada, a escolha entre um deles é feita aleatoriamente. Se o j -ésimo QTL é escolhido, então, todos os parâmetros referentes a ele são excluídos do modelo. A probabilidade de aceitação desse novo modelo é dada pelo mínimo entre 1 e A (Yi & Xu, 2001; Jannink & Fernando, 2004).

$$A = \frac{p_a}{p_e} \times \frac{p(y|\theta^*)}{p(y|\theta)} \times \frac{l}{p(U_j^m, U_j^p | \lambda_j, G_j^E, G_j^D)} \times \frac{q(U_j^m, U_j^p)}{l_0}. \quad (30)$$

Em (30), θ^* contém todos os elementos de θ , exceto aqueles correspondentes ao j -ésimo QTL. Se a proposta de excluir um QTL do modelo é aceita, o modelo no estado atual da cadeia passa a ter $(l-1)$ QTLs com vetor de parâmetros θ^* . Caso contrário, o modelo permanece com l QTLs e vetor de parâmetros θ . O esquema básico para o algoritmo com saltos reversíveis está representado na Figura 2 a seguir.

O processo de atualização dos parâmetros é repetido a cada iteração

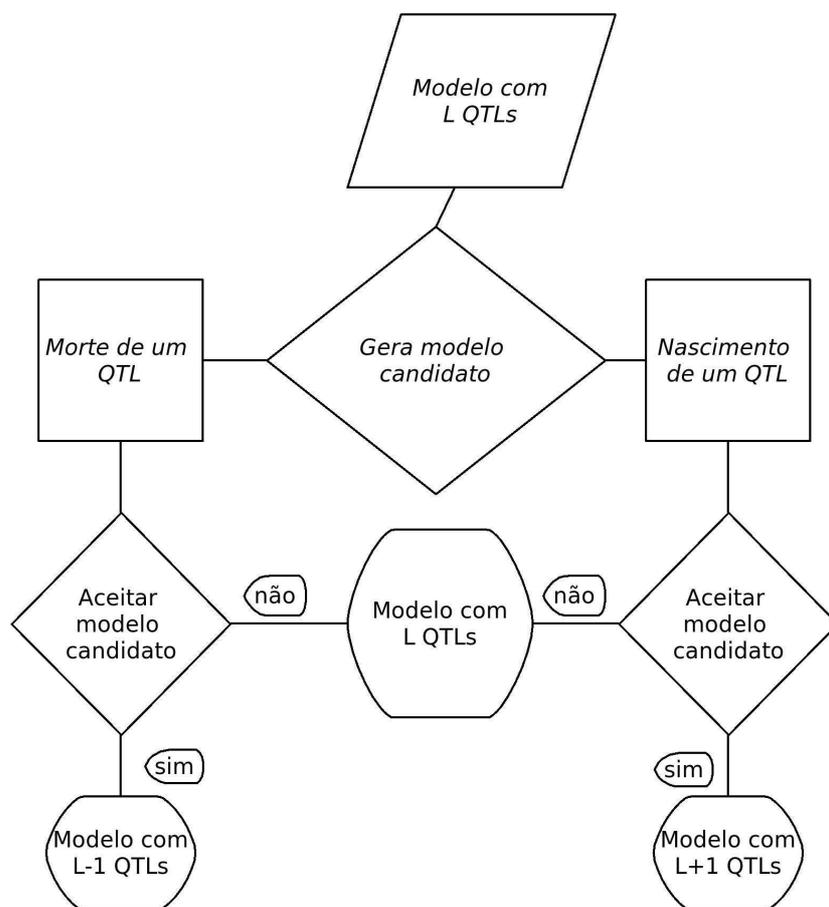


FIGURA 2: Esquema de decisões para o algoritmo com saltos reversíveis.

para todos os QTLs existentes no modelo neste ciclo, até obter-se uma amostra suficientemente grande da distribuição a posteriori desejada.

Na verificação da convergência das cadeias, foram utilizados apenas procedimentos informais de verificação gráfica, uma vez que a variação na dimensão do espaço paramétrico das cadeias dificulta a utilização dos procedimentos formais de convergência. De acordo com esses procedimentos, os comportamentos gráficos das cadeias na parte inicial e na parte final são comparados. Havendo diferenças acentuadas entre esses gráficos, assume-se

que a convergência ainda não foi atingida. Alternativamente, também são comparadas cadeias obtidas em paralelo, a partir de diferentes valores iniciais. De forma similar, se as duas cadeias apresentam gráficos com diferenças acentuadas, existem indícios de falta de convergência.

Os valores iniciais foram definidos da mesma forma para todas as análises realizadas. O número de QTLs no modelo foi inicialmente considerado igual a 3. Considerando o modelo com apenas um cromossomo, as posições iniciais para os três QTLs foram 0,125, 0,372 e 0,643M. Nos exemplos em que foram simulados dois cromossomos, as posições iniciais foram 0,125 e 0,372M no cromossomo 1 e 1,643 que corresponde à posição 0,643M do cromossomo 2.

Os efeitos aditivos e de dominância foram simulados inicialmente de distribuições normais com média zero e variâncias 0,14 e 0,10, respectivamente.

Para a média geral, variância residual, variâncias aditivas e de dominância, os valores iniciais foram iguais aos valores paramétricos.

Os valores iniciais para indicadores de segregação foram sorteados de distribuições uniformes e, a partir desses valores, foram construídas as matrizes iniciais Z e W.

4 RESULTADOS E DISCUSSÃO

4.1 Resultados da simulação 1

Considerando o conjunto de dados da estratégia 1 de simulação, os resultados das análises estão apresentados a seguir.

As probabilidades a posteriori para o parâmetro número de QTL estão apresentadas na Tabela 2, a seguir.

TABELA 2: Probabilidades a posteriori para o número de QTLs em 100.000 iterações (nf=5, nfc=6, nc=2, nQTL=3, dM=10).

| | número de QTL | | | | | |
|-------|---------------|--------|--------|--------|--------|--------|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| prob. | 0,0195 | 0,3956 | 0,4427 | 0,1277 | 0,0129 | 0,0014 |

Verifica-se, pelos dados da Tabela 2, que a densidade a posteriori estimada, para o número de QTLs, foi maior que zero para todos os possíveis valores, indicando que o algoritmo de amostragem “visitou” todos os diferentes modelos. Verifica-se, ainda, que a maior probabilidade a posteriori foi obtida para o modelo com 3 QTLs. Este resultado é um claro indício de que a abordagem apresentada permite inferir sobre este parâmetro de grande interesse e de difícil análise nos problemas de mapeamento, conforme discutido em Sillanpää & Arjas (1998).

As médias estimadas a posteriori dos parâmetros média geral (b) e variância residual (σ_e^2) estão apresentadas na Tabela 3, com seus respectivos intervalos de credibilidade.

Verifica-se pelos dados desta Tabela, que os intervalos de credibilidade para média geral e variância residual contêm os verdadeiros valores. Para a variância residual, a estimativa pontual foi abaixo do valor paramétrico.

As distribuições a posteriori aproximadas para os parâmetros média e variância residual estão apresentadas na Figura 3.

TABELA 3: Estimativas para os parâmetros média geral e variância residual em 100.000 iterações (nf=5, nfc=6, nc=2, nQTL=3, dM=10). (LI limite inferior, LS limite superior, vp valor paramétrico).

| parâmetro | média | intervalo de credibilidade | | vp |
|--------------|---------|----------------------------|-------|----|
| | | LI | LS | |
| b | - 0,053 | - 0,824 | 0,761 | 0 |
| σ_e^2 | 0,708 | 0,297 | 1,177 | 1 |

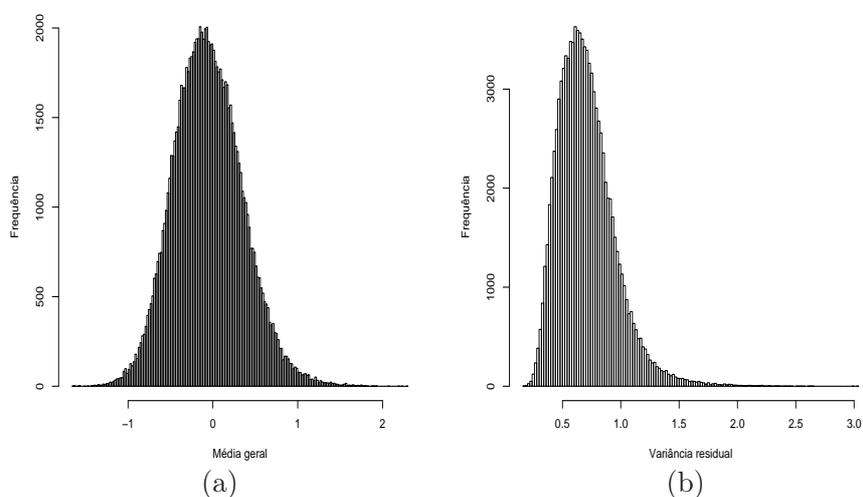


FIGURA 3: Distribuições a posteriori aproximadas para a média geral e variância residual (100.000 iterações, nf=5, nfc=6, nc=2, nQTL=3, dM=10).

Nas distribuições a posteriori da Figura 3, o valor paramétrico para a média geral encontra-se na região da moda e, para a variância residual, dentro de um intervalo com densidades altas.

Os gráficos da intensidade a posteriori de QTL, utilizada para inferir sobre a posição dos QTLs nos cromossomos, estão apresentados nas Figuras 4 e 5. De acordo com esse método, a região de maior densidade no gráfico representa a posição no cromossomo com maior probabilidade a posteri-

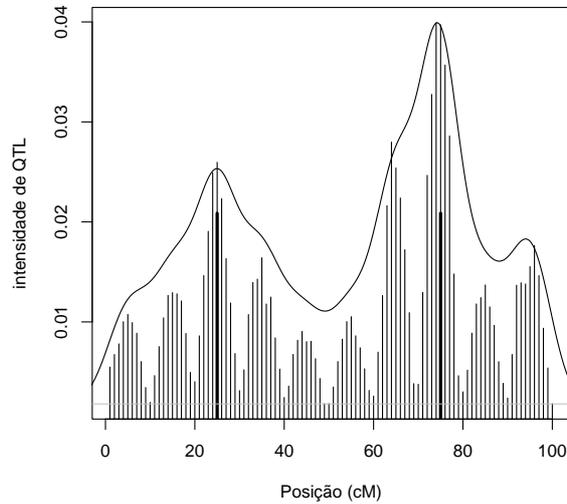


FIGURA 4: Intensidade a posteriori de QTL no cromossomo 1, com intervalos de 1cM (100.000 iterações, $nf=5$, $nfc=6$, $nc=2$, $nQTL=3$, $dM=10$).

ori de conter um QTL. As densidades apresentadas foram obtidas a partir de 10.000 reamostragens dos 100 valores de intensidade para os intervalos considerados no cromossomo 1 e 70 no cromossomo 2. Os valores das probabilidades de aceitação do algoritmo Metropolis Hastings, para as novas posições amostradas, foram de, aproximadamente, 20 por cento, podendo ser considerados satisfatórios, nesse tipo de análise. Na Figura 4 estão apresentados os gráficos da intensidade a posteriori de QTL no cromossomo 1. As posições verdadeiras dos QTLs nos cromossomos estão destacadas com linhas verticais mais escuras nos gráficos.

Pode-se verificar pela Figura 4, que existiram densidades em todas as regiões do cromossomo. A região de maior intensidade a posteriori abrange o intervalo de 60 a 80cM, sendo a maior densidade verificada entre 70 e 80cM, com moda no ponto 75cM, que corresponde à verdadeira posição

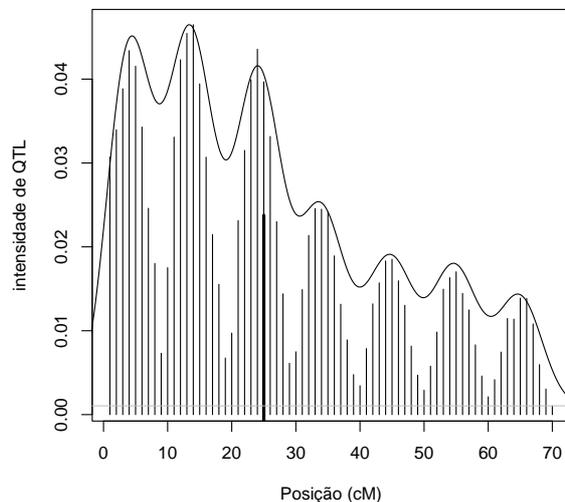


FIGURA 5: Intensidade a posteriori de QTL no cromossomo 2, com intervalos de 1cM (100.000 iterações, $nf=5$, $nfc=6$, $nc=2$, $nQTL=3$, $dM=10$).

para um dos QTLs simulados no cromossomo 1.

Uma segunda região de alta intensidade a posteriori de QTL abrange o intervalo de 20 a 30cM, com maiores densidades em 25cM, também coincidindo com o valor paramétrico para a posição do primeiro QTL simulado nesse cromossomo.

Nos resultados apresentados na Figura 4, existem evidências de que as posições possam ser estimadas corretamente, embora com baixa precisão.

A intensidade a posteriori de QTL no cromossomo 2 pode ser verificada na Figura 5.

De acordo com a Figura 5, foram obtidas densidades em todas as posições do cromossomo. A região do cromossomo 2 de maior intensidade a posteriori de QTL abrange o intervalo de 1 a 30cM, que inclui o valor paramétrico para a posição de QTL nesse cromossomo. No entanto, as

altas densidades em outras regiões dificultam a inferência sobre a posição do QTL nesse segundo cromossomo.

Sem dúvida, a principal razão para esta falta de precisão foi a utilização de um delineamento muito pequeno. No entanto, a escolha do delineamento, neste trabalho, baseou-se na limitação de esforço computacional, considerando que pedigrees com maior número de indivíduos acarretam muito tempo para a obtenção de cadeias válidas para a análise.

Uma importante questão a ser considerada nesse resultado está relacionada com a convergência das cadeias geradas pelos algoritmos de amostragem. Os critérios comumente utilizados para verificar a convergência das cadeias não puderam ser utilizados nesse método, pois eles precisariam de importantes adaptações para serem aplicados a problemas em que a dimensão do vetor de parâmetros muda de uma iteração para outra. Os critérios baseados em verificação gráfica são sempre mais subjetivos e a falta de um critério formal leva a uma suspeita de não convergência. O comportamento gráfico das três cadeias utilizadas para compor a amostra final foi semelhante para todos os parâmetros, com exceção da posição, que apresentou comportamentos ligeiramente diferentes para cada uma das cadeias. Assim, na possibilidade de utilização de um critério formal, como, por exemplo, o de Gelman e Rubin (Nogueira et al., 2004), a convergência talvez não fosse confirmada para esse parâmetro.

Para os componentes de variância referentes aos efeitos dos QTLs, as estimativas a posteriori são determinadas para cada região de maior intensidade de QTL. As distribuições a posteriori para as variâncias aditiva e de dominância, determinadas do intervalo de 20 a 30cM, e do intervalo de 70 a 80cM do cromossomo 1, estão apresentadas na Figura 6.

De acordo com a Figura 6, as distribuições a posteriori aproximadas para as variâncias aditiva e de dominância foram semelhantes para as duas regiões de altas densidades analisadas, ou seja, todas apresentaram os valores modais e médios entre 0,10 e 0,20. Assim, as diferenças entre va-

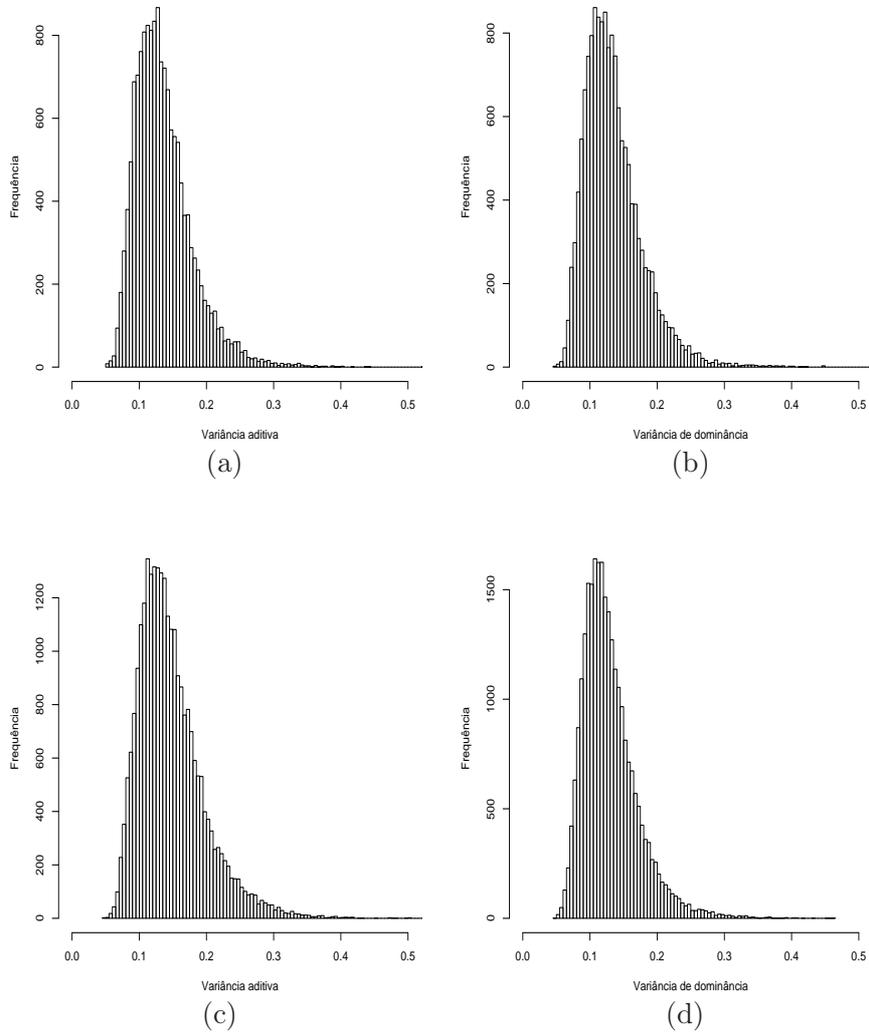


FIGURA 6: Distribuições a posteriori aproximadas das variâncias aditiva e de dominância do QTL no intervalo de 20 a 30cM (*a e b*) e no intervalo de 70 a 80cM (*c e d*) no cromossomo 1 (100.000 iterações, $nf=5$, $nfc=6$, $nc=2$, $nQTL=3$, $dM=10$).

riâncias genéticas dos QTLs, estimados em posições diferentes, não foram detectadas nesta análise.

Em delineamentos muito pequenos, a variação genética presente nos dados é pouco informativa, permitindo estimar precariamente principalmente as componentes da variância. Uma alternativa nesses casos é o uso de prioris mais informativas para os componentes de variância, propiciando, assim, inferências viáveis para os demais parâmetros de interesse. Em contrapartida, as distribuições a posteriori para essas variâncias são penalizadas, sofrendo influência maior das distribuições a priori.

Os resultados apresentados, embora não estimando precisamente todas as posições, dão indícios de que a utilização do pedigree completo mostra poder de detecção de QTL, conforme já discutido por Yi & Xu (2001). A estimativa correta do número de QTLs a influenciar a característica, para o primeiro delineamento considerado, evidencia o fato de que o método empregado pode estimar o modelo mais adequado para os dados, mesmo utilizando um pedigree com tamanho reduzido.

4.1.1 Análise condicional ao número de QTLs

Em análises alternativas, considerou-se o modelo com um número fixo de QTLs, objetivando verificar se, na ausência do algoritmo com saltos reversíveis, a posição seria melhor estimada, uma vez que a seleção do modelo nesse caso é momentaneamente ignorada. Apresenta-se, a seguir, apenas a análise do modelo considerando 3 QTLs, sendo descartadas as 1.000 iterações iniciais e utilizado um salto de 10 iterações, obtendo-se no final, cadeias com 50.000 amostras válidas.

As distribuições a posteriori para os parâmetros média geral e variância residual estão apresentadas na Figura 7.

Na Figura 7 verifica-se que a média geral foi estimada corretamente e, da mesma forma que na análise anterior, a moda da distribuição a posteriori para a variância residual apresentou um valor abaixo do valor paramétrico.

As distribuições a posteriori para as posições dos QTLs, nos dois cromossomos, estão apresentadas na Figura 8, de acordo com o método

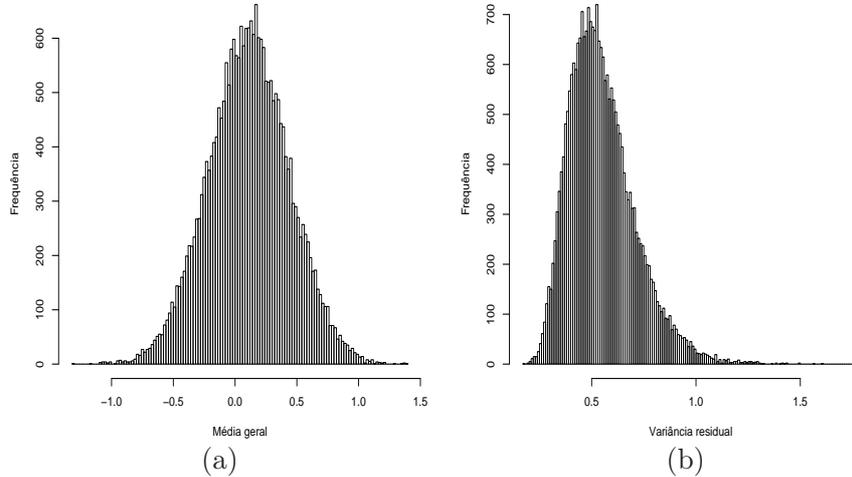


FIGURA 7: Distribuições a posteriori aproximadas para a média geral e variância residual (50.000 iterações, $nf=5$, $nfc=6$, $nc=2$, $nQTL=3$, $dM=10$, modelo fixo em 3 QTLs).

da intensidade a posteriori de QTL. As linhas verticais mais escuras nos gráficos correspondem às verdadeiras posições dos QTLs.

As probabilidades de aceitação do algoritmo Metropolis Hastings para as novas posições amostradas ficaram em torno de 10 por cento.

No cromossomo 1, de acordo com a Figura 8 (a), as maiores intensidades de QTL estão entre as posições 60 e 80cM, sendo a maior evidência de QTL encontrada entre as posições 70 e 80. Assim, o QTL simulado em 75cM do cromossomo 1 teve sua posição identificada corretamente. No entanto, nesse mesmo cromossomo, o QTL simulado em 25cM, não teve sua posição identificada, ou seja, as densidades foram baixas para se inferir sobre a existência de QTL nessa posição.

No cromossomo 2, Figura 8 (b), várias regiões apresentaram valores altos de intensidade de QTLs, sendo a maior delas correspondente ao intervalo entre 0 e 7cM. Assim, seria mais provável afirmar que existe um QTL nesse intervalo, embora a posição 25cM, correspondente à verdadeira posi-

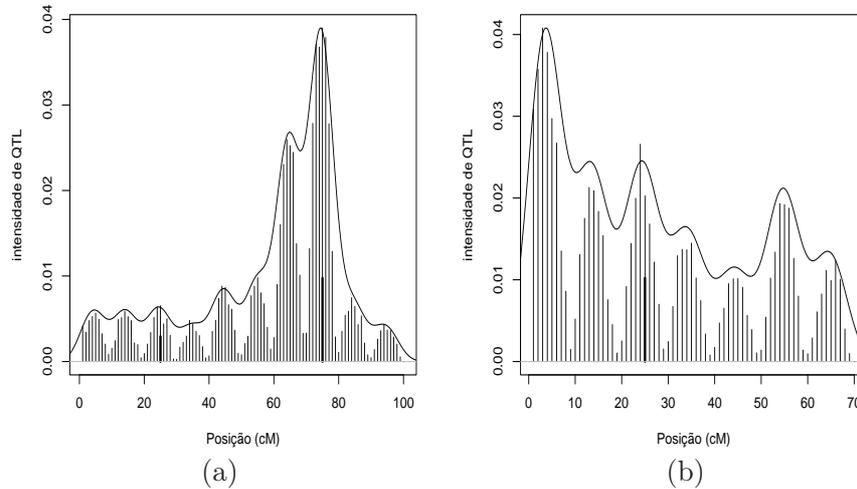


FIGURA 8: Intensidade a posteriori de QTL nos cromossomos 1 e 2, com intervalos de 1cM (50.000 iterações, $nf=5$, $nfc=6$, $nc=2$, $nQTL=3$, $dM=10$, modelo fixo em 3 QTLs).

ção de QTL nesse cromossomo, também tenha apresentado valores altos de intensidade.

Este resultado permite estimar precisamente apenas uma das posições verdadeiras de QTL no cromossomo 1. Portanto, ao contrário do que seria de se esperar, fixar o número correto de QTLs não conduziu à inferência condicional mais eficiente das posições para os QTLs. Pode-se especular que a convergência ainda não tenha sido atingida, podendo essa ser a causa da não estimação correta das outras posições. No entanto, o delineamento pequeno segue sendo a explicação mais plausível para a má qualidade da inferência neste caso.

Nos gráficos da intensidade a posteriori, foram verificados picos de densidades nas posições múltiplas de 5, com exceção daquelas múltiplas de 10, nas quais, ao contrário, foram verificados os menores valores de intensidade. Assim, em posições próximas ou coincidentes com os marcadores, identificou-se que, provavelmente, não existem QTLs. Em consequência

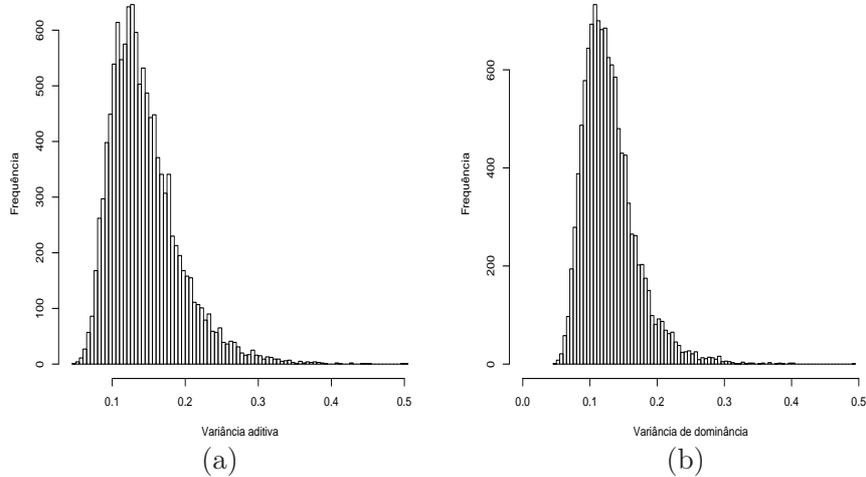


FIGURA 9: Distribuições a posteriori aproximadas para as variâncias aditiva e de dominância no intervalo de 70 a 80cM no cromossomo 1 (50.000 iterações, $nf=5$, $nfc=6$, $nc=2$, $nQTL=3$, $dM=10$, modelo fixo em 3 QTLs).

disso, no conjunto de dados 2, foram simulados valores diferentes para as posições, com o objetivo de verificar se esse comportamento gráfico foi influenciado pela simulação de valores múltiplos de 5.

As variâncias aditiva e de dominância do QTL estimado no cromossomo 1 estão apresentadas na Figura 9. Nesta Figura, pode-se verificar que a moda da distribuição a posteriori apresentou valores maiores que verdadeira variância de dominância, simulada igual a 0,05. Para a variância aditiva, a moda da distribuição encontra-se bem próxima do valor simulado igual a 0,10. Pode-se supor que o primeiro QTL localizado nesse mesmo cromossomo, e cuja posição não foi detectada pelo método, tenha influenciado na estimação das variâncias.

De maneira geral, a análise do modelo com número fixo de QTLs, mesmo ignorando o problema da seleção de modelos, não apresentou resultados comparativamente melhores que aqueles do RJ, nem mesmo consi-

derando o número verdadeiro de QTLs simulados. Conforme discutido em muitos trabalhos com esse enfoque, espera-se um melhor desempenho da análise com saltos reversíveis, embora, em trabalhos mais recentes, já se discutam alternativas de mapeamento que possam superar ou aperfeiçoar essa técnica, considerando, por exemplo, cromossomos supersaturados de QTLs. Assim, tem-se que este resultado foi coerente com os demais trabalhos da literatura, os quais afirmam que o modelo saturado de QTLs ou o RJ apresentam resultados melhores do que aqueles fixando em modelos com um pequeno número de QTLs.

4.2 Resultados da simulação 2

De acordo com as análises dos dados obtidos na segunda estratégia de simulação, a distribuição a posteriori para o número de QTLs a ser considerado no modelo está apresentada na Tabela 4.

TABELA 4: Probabilidades a posteriori para o número de QTL em 9.000 amostras válidas ($n_f=5$, $n_{fc}=10$, $n_c=1$, $n_{QTL}=2$, $dM=10$).

| | número de QTL | | | | | |
|-------|---------------|-------|-------|-------|-------|-------|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| prob. | 0,534 | 0,359 | 0,096 | 0,011 | 0,000 | 0,000 |

O algoritmo Metropolis Hastings com saltos reversíveis oscilou entre os modelos com número de QTLs de 1 até 4, não havendo densidades não nulas para os modelos com 5 e 6, de acordo com a Tabela 4. O modelo com 2 QTLs apresentou uma alta densidade a posteriori, mas a maior probabilidade foi registrada para o número de QTLs igual a 1. Assim, os resultados indicam que o modelo mais adequado deve considerar apenas 1 QTL.

As distribuições a posteriori aproximadas para os parâmetros média e variância residual estão apresentadas na Figura 10.

De acordo com as distribuições na Figura 10, o valor simulado igual a 1 para a variância não apresentou altas densidades. Para a média geral,

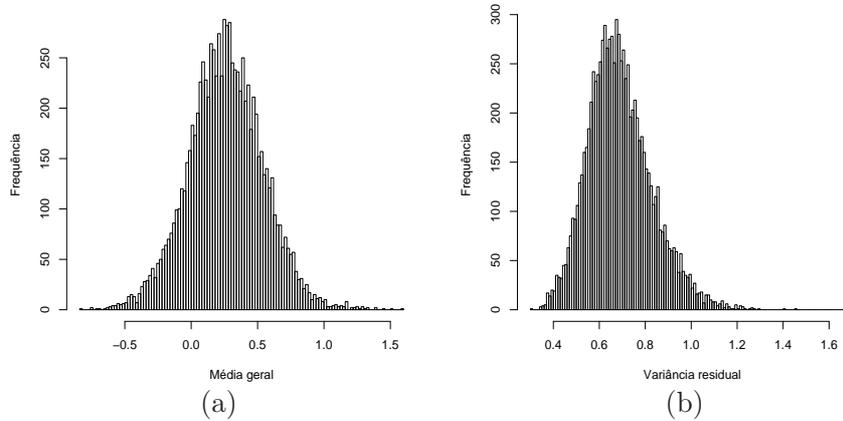


FIGURA 10: Distribuições a posteriori aproximadas para a média geral e variância residual. (9.000 iterações, $nf=5$, $nfc=10$, $nc=1$, $nQTL=2$, $dM=10$).

os valores de altas densidades foram mais próximos do valor verdadeiro.

As médias estimadas a posteriori dos parâmetros média geral (b) e variância ambiental (σ_e^2) estão apresentadas na Tabela 5, com seus respectivos intervalos de credibilidade.

TABELA 5: Estimativas para os parâmetros média geral e variância residual em 9.000 iterações ($nf=5$, $nfc=10$, $nc=1$, $nQTL=2$, $dM=10$). (LI limite inferior, LS limite superior, vp valor paramétrico).

| parâmetro | média | intervalo de credibilidade | | vp |
|--------------|-------|----------------------------|-------|----|
| | | LI | LS | |
| b | 0,261 | - 0,315 | 1,791 | 0 |
| σ_e^2 | 0,689 | 0,440 | 0,985 | 1 |

Pelos dados da Tabela 5, verifica-se que o intervalo de credibilidade contém o verdadeiro valor apenas para a média geral, não sendo, portanto, estimada corretamente a variância residual.

A posição estimada do QTL no cromossomo pode ser verificada por

meio da intensidade a posteriori do QTL apresentada na Figura 11. As taxas de aceitação para os novos valores de posições, gerados de acordo com o Metropolis Hastings, ficaram em torno de 10 por cento.

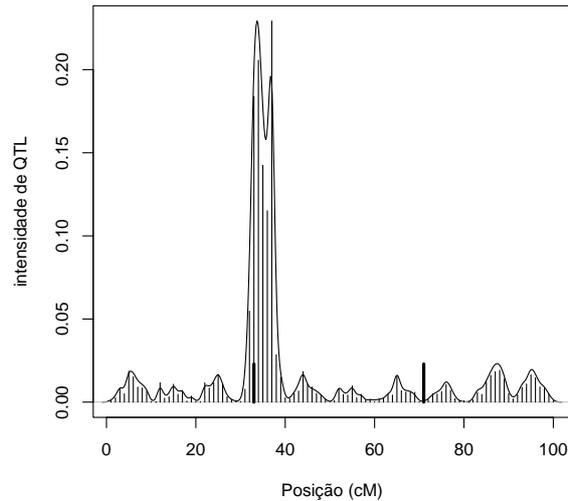


FIGURA 11: Intensidade a posteriori de QTL com intervalos de 1cM no cromossomo 1. (9.000 iterações, $nf=5$, $nfc=10$, $nc=1$, $nQTL=2$, $dM=10$).

De acordo com a Figura 11, a região de maior intensidade de QTL compreende o intervalo de 32 a 38cM. A posição do primeiro QTL simulado nesse cromossomo (34cM) encontra-se exatamente na moda da distribuição, sendo possível afirmar que foi identificado um QTL nessa posição. As demais regiões do cromossomo não apresentaram valores altos de intensidade, inclusive a posição do segundo QTL simulado no cromossomo e não detectado pelo algoritmo com saltos reversíveis. Assim, verificou-se, nessa análise, uma coerência no resultado da estimação do número de QTLs com a posição, que apresentou apenas uma região de alta densidade.

Comparando-se os resultados com aqueles obtidos na análise da si-

mulação 1, nota-se que, embora um QTL não tenha sido detectado, a estimação da posição do QTL mostrou-se mais eficiente e precisa, com altas intensidades de QTL apenas na região da posição correta. O aumento do número de indivíduos no pedigree e a redução para apenas 1 cromossomo são fatores que, provavelmente, influenciaram na melhor identificação da posição. Verificou-se também que o aumento do número de indivíduos no pedigree acarretou uma diminuição nas probabilidades de aceitação tanto do número de QTLs como da posição, indicando que seria necessário um número maior de iterações para se obter uma distribuição a posteriori com uma mistura mais adequada para a análise.

As distribuições a posteriori estimadas para as variâncias aditivas e de dominância estão apresentadas na Figura 12, considerando apenas o região do cromossomo que apresentou maior intensidade a posteriori de QTL.

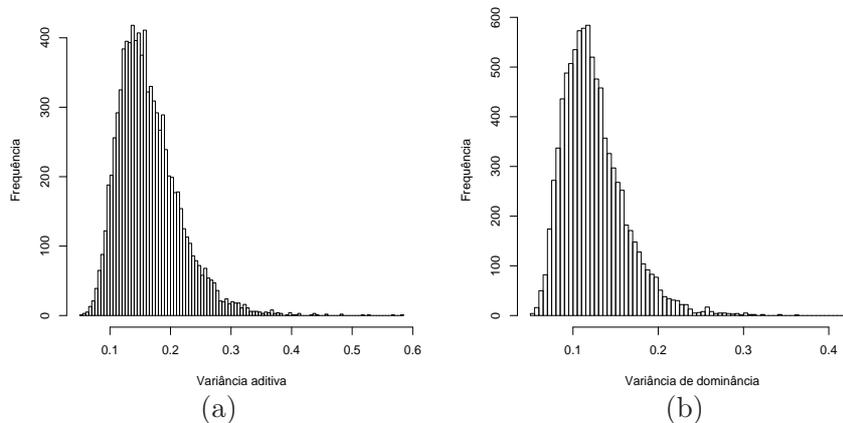


FIGURA 12: Distribuições a posteriori aproximadas das variâncias aditiva e de dominância do QTL no intervalo de 30 a 40cM no cromossomo 1. (9.000 iterações, $nf=5$, $nfc=10$, $nc=1$, $nQTL=2$, $dM=10$).

Considerando que o valor simulado para a variância aditiva do QTL

localizado na posição 34cM é de 0,15, nota-se, na Figura 12 (a), que a moda da distribuição a posteriori concentra-se em torno desse valor. Para a distribuição a posteriori da variância de dominância, a moda encontra-se mais próxima de 0,10, que corresponde ao valor simulado para este parâmetro.

Na análise desse segundo conjunto de dados, verifica-se que, em geral, foram obtidos melhores resultados, podendo-se concluir que inferências válidas são possíveis com esse método.

As estimativas corretas obtidas com esse método reforçam a hipótese de que, utilizando-se um delineamento melhor, seja possível estimar mais precisamente todos os parâmetros propostos nesse método.

4.3 Resultados da simulação 3

Considerando a análise da terceira estratégia de simulação, a distribuição a posteriori para o número de QTLs no modelo, obtida por meio do algoritmo RJ, está apresentada na Tabela 6.

TABELA 6: Probabilidades a posteriori para o número de QTLs em 24.0000 amostras válidas (nf=5, nfc=10, nc=1, nQTL=2, dM=2).

| | número de QTL | | | | | |
|-------|---------------|--------|--------|--------|-------|-------|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| prob. | 0,1142 | 0,6512 | 0,2131 | 0,0215 | 0,000 | 0,000 |

Verifica-se pelos dados da Tabela 6, que o número de QTLs igual a dois apresentou uma probabilidade a posteriori muito superior aos demais valores. Além disso, os números de QTLs de 1 a 4 foram todos “visitados” pelo algoritmo, indicando boa mistura dos valores amostrados. Este resultado evidencia a eficiência do algoritmo RJ na inferência sobre o modelo mais adequado, para esta situação de delineamento considerada.

As distribuições a posteriori aproximadas para os parâmetros média e variância residual estão apresentadas na Figura 13. As distribuições nesta Figura apresentaram altas densidades para o valor paramétrico da média

geral, mas, para a variância residual, a moda da distribuição foi inferior ao valor paramétrico.

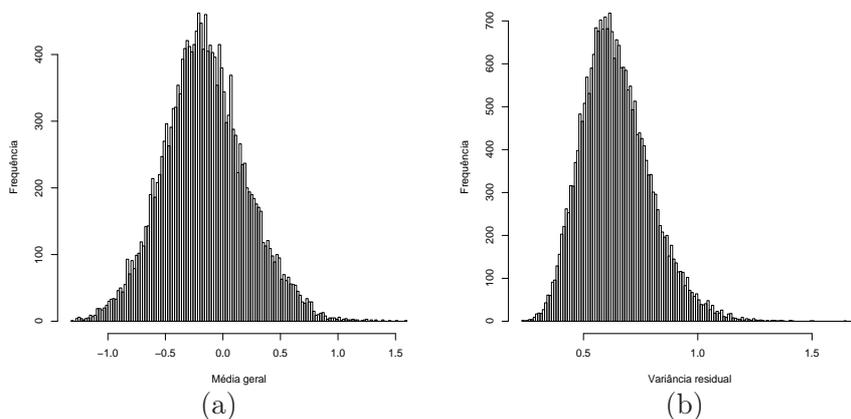


FIGURA 13: Distribuições a posteriori aproximadas para a média geral e variância residual (24.000 iterações, $nf=5$, $nfc=10$, $nc=1$, $nQTL=2$, $dM=2$).

O gráfico da intensidade a posteriori de QTL está apresentado na Figura 14, com as verdadeiras posições destacadas em linhas verticais mais escuras.

Na Figura 14, percebem-se duas regiões de maiores intensidades que podem sugerir a existência de QTLs a elas associados. A primeira delas inclui o intervalo de 30 a 40cM e a segunda, o intervalo de 65 a 85cM, ambas contendo as posições verdadeiras para os QTLs simulados nesse cromossomo. Este resultado sugere que as estimativas apresentam uma tendência a convergir para os valores paramétricos das posições. No entanto, houve pouca mistura nos valores amostrados e o número de iterações e o de salto não foram suficientes para se obter a convergência, conforme pode ser visto claramente na Figura 14.

Segundo Wang et al. (2005), a dificuldade na convergência das cadeias e a fraca mistura dos algoritmos de amostragem têm sido notadas

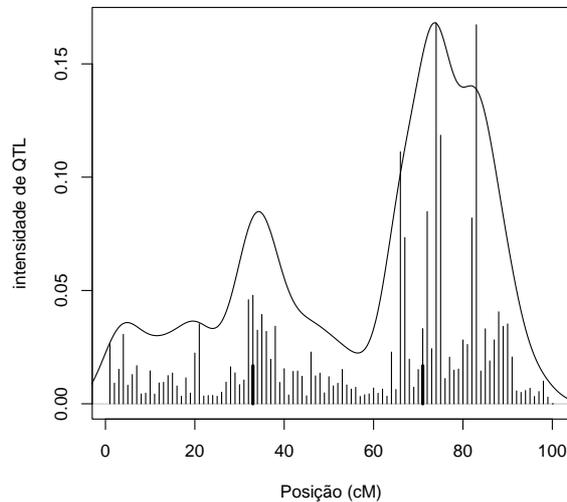


FIGURA 14: Intensidade a posteriori de QTL com intervalos de 1cM no cromossomo 1 (24.000 iterações, $nf=5$, $nfc=10$, $nc=1$, $nQTL=2$, $dM=2$).

frequentemente em trabalhos utilizando o RJ.

As distribuições a posteriori estimadas para as variâncias aditivas e de dominância estão apresentadas na Figura 15, considerando apenas as regiões do cromossomo que apresentaram maior intensidade a posteriori de QTL.

De acordo com a Figura 15, também nessa análise foram obtidas distribuições a posteriori semelhantes para as variâncias genéticas dos dois QTLs simulados em posições distintas do cromossomo. Uma possível razão para este fato é o uso de uma distribuição a priori muito informativa para componentes de variância, em relação ao conjunto de dados, que é muito pequeno para fornecer informação mais precisa sobre essas variâncias. Assim, a informação a priori é mais influente sobre a posteriori do que a informação dos dados. No anexo B, podem-se comparar os gráficos das

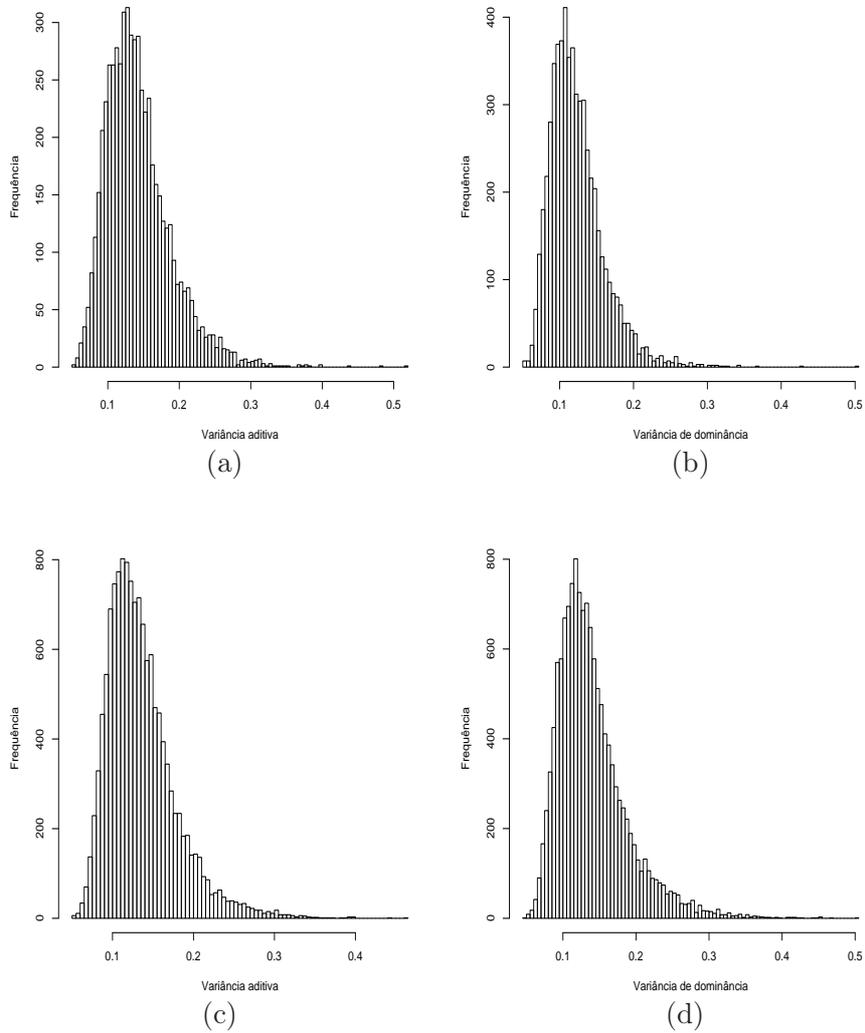


FIGURA 15: Distribuições a posteriori aproximadas das variâncias aditiva e de dominância no intervalo de 30 a 40cM (*a* e *b*) e no intervalo de 65 a 85cM (*c* e *d*) no cromossomo 1. (24.000 iterações, $nf=5$, $nfc=10$, $nc=1$, $nQTL=2$, $dM=2$).

distribuições a priori e das distribuições a posteriori para componentes de variância, plotados em uma mesma figura.

O aumento do número de marcadores por cromossomo e do número de iterações permitiu estimar com maior precisão o número de QTLs no modelo. Com relação à posição, esperava-se uma maior eficiência do método na estimação, comparando-se com os resultados da simulação 2. No entanto, verificaram-se apenas indícios de que as duas posições possam ser estimadas corretamente. Ao que tudo indica, melhores resultados são esperados para um aumento significativo no tamanho do pedigree. Outra alternativa que também poderia ter sido considerada é a seleção prévia de marcadores conforme a análise de marcas individuais, eliminando-se aquelas pouco informativas ou correlacionadas, como forma de melhorar a informação dos marcadores sobre a posição dos QTLs.

Ao longo de todo o trabalho, um dos maiores problemas encontrados foi a otimização de programas. O esforço computacional é grande e a maior preocupação foi com a correção da metodologia e não com a eficiência da programação. Isto tornou-se uma limitação da técnica para problemas e delineamentos mais realistas ($nf > 20$, $nfc > 50$).

É preciso otimizar as técnicas de programação para possibilitar a obtenção de resultados mais conclusivos. Isso depende do envolvimento e da interação entre diversos grupos de pesquisa interessados em mapear QTLs em pedigrees complexos.

Além da implementação de uma programação mais eficiente, seria de grande interesse, em trabalhos posteriores, a criação de um software que possa disponibilizar a metodologia para pesquisadores que, na prática, necessitem analisar problemas semelhantes. No caso em estudo, seria importante disponibilizar bibliotecas R para análise bayesiana de QTLs em pedigrees complexos, a exemplo do que já existe para delineamentos simples.

5 CONCLUSÕES

A implementação do algoritmo de amostragem com saltos reversíveis mostrou-se uma alternativa eficiente, permitindo inferir adequadamente sobre o número de QTLs para a maioria das situações analisadas.

A inferência condicional ao número de QTLs, por si só, não foi suficiente para propiciar estimativas mais precisas para os demais parâmetros, especialmente para as posições.

Mapeamento de QTLs em pedigrees complexos exigem técnicas mais eficientes de programação que permitam explorar todo o poder do método com a utilização de delineamentos mais realistas em relação ao número de indivíduos.

REFERÊNCIAS BIBLIOGRÁFICAS

- ALMASY, L.; BLANGERO, J. Multipoint quantitative-trait linkage analysis in general pedigree. **The American Journal of Human Genetics**, Chicago, v. 62, p. 1198-1211, May 1998.
- BERGER, J. **Statistical decision theory and Bayesian analysis**. New York: Springer-Verlag, 1985.
- BERNARDO, J.M.; SMITH, A.F.M. **Bayesian theory**. New York: J. Wiley, 1994.
- BINK, M.C.A.M.; JANSSE, L.L.G.; QUASS, R.L. Markov chain Monte Carlo for mapping a quantitative trait locus in outbred populations. **Genetical Research**, New York, v. 75, n. 2, p. 231-241, Apr. 2000.
- BINK, M.C.A.M.; ARENDONK, A.M. van. Detection of quantitative trait loci in outbred population with incomplete marker data. **Genetics**, Baltimore, v. 151, n. 1, p. 409-420, Jan. 1999.
- BOX, G.E.; TIAO, G.C. **Bayesian inference in statistical analysis**. New York: J. Wiley, 1992. 588 p.
- CHIB, S.; GREENBERG, E. Understanding the Metropolis-Hastings algorithm. **The American Statistician**, Alexandria, v. 49, n. 4, p. 327-335, Dec. 1995.
- 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 genes action. **Genetics**, Baltimore, v. 116, n. 1, p. 113-125, May 1987.
- EHLERS, R.S. **Métodos computacionalmente intensivos em estatística**. 2004. 38 p. Disponível em: <<http://leg.ufpr.br/~ehlers/notas/mci.pdf>>. Acesso em: 2 jun. 2008.
- ELSTON, R.C.; STEWART, J. The analysis of quantitative traits for simple genetic models from parental F1 and backcrosses data. **Genetics**, Baltimore, v. 73, n. 4, p. 695-711, 1973.

- FERREIRA, M.E.; GRATTAPAGLIA, D. **Introdução ao uso de marcadores RAPD, RFLP em análise genética**. Brasília: EMBRAPA - CENARGEN, 1995. 220 p.
- GELDERMANN, H. Investigations on inheritance of quantitative characteres in animal by gene markers. **Theoretical and Applied Genetics**, Berlin, v. 46, n. 7, p. 319-330, Dec. 1975.
- GELFAND A.E. Gibbs sampling. **Journal of the American Statistical Association**, Alexandria, v. 95, n. 452, p. 1300-1304, 2000.
- GELFAND, A.E.; SMITH A.F.M. Sampling-based approaches to calculating marginal densities. **Journal of the American Statistical Association**, Alexandria, v. 85, n. 410, p. 348-409, June 1990.
- GELMAN, A.; CARLIN, J.B.; STERN, H.S.; RUBIN, D.B. **Bayesian data analysis**. London: Chapman and Hall, 1997. 526 p.
- GEMAN, S. GEMAN, D. Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, New York, v. 6, n. 6, p. 721-741, 1984.
- GEORGE, A.W.; VISSCHER, P.M.; HALEY, C.S. Mapping quantitative trait loci in complex pedigrees: a two-step variance component approach. **Genetics**, Baltimore, v. 156, n. 4, p. 2081-2092, Dec. 2000.
- GOFFINET, B.; GERBER, S. Quantitative trait loci: a meta-analysis. **Genetics**, Baltimore, v. 155, n. 1, p. 463-473, May 2000.
- GREEN, P.J. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. **Biometrika**, London, v. 82, n. 4, p. 711-732, Dec. 1995.
- 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.
- HALEY, C.S.; KNOTT, S.A. A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. **Heredity**, Oxford, v. 69, n. 4, p. 315-324, Oct. 1992.

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

HEATH S.C. Markov chain Monte Carlo segregation and linkage analysis for oligogenic models. **The American Journal of Human Genetics**, Chicago, v. 61, n. 3, p. 748-760, Sept. 1997.

JANNINK, J.L.; FERNANDO, R. On the Metropolis-Hastings acceptance probability to add or drop a quantitative trait locus in Markov chain Monte Carlo based Bayesian analyses. **Genetics**, Baltimore, v. 166, n. 1, p. 641-643, Jan. 2004.

JANSEN, R.C. A general mixture model for mapping quantitative trait loci by using molecular markers. **Theoretical and Applied Genetics**, Berlin, v. 85, n. 2/3, p. 252-260, Nov. 1992.

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

JANSEN, R.C.; JOHNSON, D.L.; ARENDONK, J.A.M. van. A mixture model approach to the mapping quantitative trait loci in complex populations with an application to multiple cattle families. **Genetics**, Baltimore, v. 148, n. 1, p. 391-399, Jan. 1998.

JEFFREYS, H. **Theory of probability**. Oxford: Clarendon Press, 1961. 447 p.

JIANG, C.; ZENG, Z.B. Mapping quantitative trait loci with dominant and missing markers in various crosses from two inbred lines. **Genetics**, Baltimore, v. 101, n. 1, p. 47-58, 1997.

JIANG, C.; ZENG, Z.B. Multiple trait analysis of genetic mapping for quantitative trait loci. **Genetics**, Baltimore, v. 140, n. 3, p. 1111-1127, July 1995.

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, July 1999.

KASS, R.E.; RAFTERY, A.E. Bayes factors. **Journal of the American Statistical Association**, Alexandria, v. 90, n. 430, p. 773-795, June 1995.

KNOTT, S.A.; HALEY, C.S. Aspects of maximum likelihood methods for the mapping of quantitative trait loci in line crosses. **Genetic Research**, New York, v. 60, n. 2, p. 139-151, Oct. 1992.

KOSAMBI, D.D. The estimation of map distances from recombination values. **Annals of Eugenics**, London, v.12, 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, Jan. 1989.

LIU, B.H. **Statistical genomics: linkage, mapping and QTL analysis**. Boca Raton: CRC Press, 1998. 611 p.

LYNCH, M.; WALSH, B. **Genetics and analysis of quantitative traits**. Sunderland: Sinauer Associates, 1998. 980 p.

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

METROPOLIS, N.; ROSEMBLUT, A.W.; ROSEMBLUT, M.N.; TELLER, A.H.; TELLER, E. Equations of state calculations by fast computing machines. **Journal of Chemical Physics**, New York, v. 21, 1087-1092, 1953.

NOGUEIRA, D.A.; SÁFADI, T; FERREIRA, D.F. Avaliação de critérios de convergência para o método de Monte de Carlo via Cadeias de Markov. **Revista Brasileira de Estatística**, Rio de Janeiro, v. 65, n. 224, p. 59-88, 2004.

PAULINO, D.C.; TURKAMAN, A.A.; MURTEIRA, B. **Estatística bayesiana**. Lisboa: Fundação Calouste Gulbenkian, 2003. 447 p.

R FOUNDATION FOR STATISTICAL COMPUTING. R Development Core Team. **R: a language and environment for statistical computing**. Vienna, 2008. Disponível em: <<http://www.R-project.org>>. Acesso em: 10 out. 2008.

RODOLPHE, F.; LEFORT, M. A multi-marker model for detecting chromosomal segments displaying QTL activity. **Genetics**, Baltimore, v. 134, n. 4, p. 1277-1288, Aug. 1993.

- SATAGOPAN, J.M.; YANDELL, B.S. **Estimating the number of quantitative trait loci via Bayesian model determination**. Special contributed paper session on genetic analysis of quantitative traits and complex diseases. Biometric section, Statistical Meeting, Chicago. 1996.
- SATAGOPAN, J.M.; YANDELL, B.S.; NEWTON, M.A.; OSBORN, T.C. A Bayesian approach to detect quantitative trait loci using Markov Chain Monte Carlo. **Genetics**, Baltimore, v. 144, n. 2, p. 805-816, Oct. 1996.
- SAX, K. The association of size differences with seed-coat pattern and pigmentation in *Phaseolus vulgaris*. **Genetics**, Baltimore, v. 8, n. 2, p. 552- 560, 1923.
- SILLANPÄÄ, M.J.; ARJAS, E. Bayesian mapping of multiple quantitative trait loci from incomplete inbred line cross data. **Genetics**, Baltimore, v. 148, n. 3, p. 1373-1388, Mar. 1998.
- SILLANPÄÄ, M.J.; ARJAS, E. Bayesian mapping of multiple quantitative trait loci from incomplete outbred off spring data. **Genetics**, Baltimore, v. 151, n. 4, p. 1605-1619, Apr. 1999.
- SILVA, H.D.; VENCOVSKY, R. Poder de detecção de "quantitative trait loci", da análise de marcas simples e da regressão linear múltipla. **Scientia Agricola**, Piracicaba, v. 59, n. 4, p. 755-762, jul./ago. 2002.
- SOLLER, M.; BRODY, T.; GENIZI, A. On the power of experimental designs for the detection of linkage between marker loci and quantitative loci in crosses between inbred lines. **Theoretical and Applied Genetics**, Berlin, v. 47, n. 1, p. 35-39, 1976.
- SOLLER, M.; GENIZI, A. The efficiency of experimental designs for the detection of linkage between a marker locus and a locus affecting a quantitative trait in segregating populations. **Biometrics**, v. 34, n. 1, p. 47-55, 1978.
- SORENSEN, D.; GIANOLA, D. **Likelihood, Bayesian and MCMC methods in quantitative genetics**. New York: Springer-Verlag, 2002. 740 p.
- STEPHENS, D.A.; FISCH, R.D. Bayesian analysis of quantitative trait locus data using reversible jump Markov chain Monte Carlo. **Biometrics**, Washington, v. 54, n. 4, p. 1334-1347, Dec. 1998.

STEWART, J.; ELSTON, R.C. Biometrical genetics with one or two loci: the inheritance of physiological characters in mice. **Genetics**, Baltimore, v. 73, n. 4, p. 675-693, 1973.

TANKSLEY, S.D.; MEDINA FILHO, H.; RICK, C.M. Use of naturally occurring enzyme variation to detect and map genes controlling quantitative traits in an interspecific backcross of tomato. **Heredity**, Oxford, v. 49, n. 1, p. 11-25, Aug. 1982.

UIMARI, P.; HOESCHELE, I. Mapping-linked quantitative trait loci using Bayesian analysis and Markov chain Monte Carlo algorithms. **Genetics**, Baltimore, v. 146, n. 2, p. 735-743, June 1997.

WANG, H.; ZHANG, Y.M.; LI, X.; MASINDE, G.L.; MOHAN, S.; BAYLINK, D.J.; XU, S. Bayesian shrinkage estimation of quantitative trait loci parameters. **Genetics**, Baltimore, v. 170, n. 1, p. 465-480, May 2005.

WELLER, J.I. Maximum likelihood techniques for the mapping and the analysis of quantitative trait loci with the aid of genetic markers using approximate likelihood methods. **Biometrics**, Washington, v. 42, n. 3, p. 627-640, Sept. 1986.

WELLER, J.I.; SOLLER, M.; BRODY, T. Linkage analysis of quantitative traits in an interspecific cross of tomato (*Lycopersicon esculentum* X *Lycopersicon pimpinellifolium*) by means of genetic markers. **Genetics**, Baltimore, v. 118, n. 2, p. 329-339, Feb. 1988.

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

XU, S.; YI, N. Mixed model analysis of quantitative trait loci. **Proceedings of the National Academic of Science of the United State of America**, Washington, v. 97, n. 26, p. 14542-14547, Dec. 2000.

YI, N. A unified Markov chain Monte Carlo framework for mapping multiple quantitative trait loci. **Genetics**, Baltimore, v. 167, n. 2, p. 967-975, June 2004.

YI, N.; XU, S. Bayesian mapping of quantitative trait loci under complicated mating designs. **Genetics**, Baltimore, v. 157, n. 4, p. 1759-1771, Apr. 2001.

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, Dec. 1993.

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

ANEXOS

Anexo A. Programas utilizados na simulação e amostragem das distribuições a posteriori. 78

Anexo B. Comparações entre as distribuições a priori e a posteriori para componentes de variância. 102

Anexo A. Programas utilizados na simulação e amostragem das distribuições a posteriori.

Programa 1: Rotina R para simulação do pedigree complexo e dos valores fenotípicos.

```
##### Simulação do pedigree
m <- 5          ## número de fundadores
nfc <- 6        ## número de filhos por cruzamento
nF1 <- m*nfc
nF2 <- m*nfc
n <- m+nF1+nF2
FO <- matrix(sample(1:6,(m*44),replace=TRUE),m,44)
##### 44 = 2*22 alelos; 19 marcadores + 3 QTLs
for (i in 1:m)
{
  FO[i, 4] <- i
  FO[i,10] <- i
  FO[i,17] <- i
  FO[i,26] <- m + i
  FO[i,32] <- m + i
  FO[i,39] <- m + i
}
FOm <- FO[1:m, 1:22]
FOp <- FO[1:m,23:44]
##### Gerando estrutura de m cruzamentos F1
cF1 <- matrix(0,m,2)
cF1[,1] <- sample(1:m, m,replace=TRUE)
cF1[,2] <- sample(1:m, m,replace=TRUE)
F1m <- matrix(0,(m*nfc),22)
F1p <- matrix(0,(m*nfc),22)
np1m <- 0      ##### numero de permutas no cromossomo 1 materno
pp1m <- 0      ##### posição de permutas no cromossomo 1 materno
cr1m <- 0      ##### cromossomo 1 materno
np2m <- 0      ##### numero de permutas no cromossomo 2 materno
pp2m <- 0      ##### posição de permutas no cromossomo 2 materno
cr2m <- 0      ##### cromossomo 2 materno
np1p <- 0      ##### numero de permutas no cromossomo 1 paterno
pp1p <- 0      ##### posição de permutas no cromossomo 1 paterno
cr1p <- 0      ##### cromossomo 1 paterno
np2p <- 0
pp2p <- 0
cr2p <- 0      ##### cromossomo 2 paterno
pm<-c(0,10,20,25,30,40,50,60,70,75,80,90,100,0,10,20,25,30,40,50,60,70)
indrecm <- c(rep(0,22))
indrecp <- indrecm
gmmae <- 0*FOm[,1]
gmpai <- 0*FOp[,1]
gmat <- cbind(gmmae,gmpai)
gmatr <- gmat
gpmae <- 0*FOm[,1]
gppai <- 0*FOp[,1]
gpat <- cbind(gpmae,gppai)
gpatr <- gpat
```

```

pais      <- c(rep(0,m))
maes     <- c(rep(0,m))
individuo <- 0
for(cross in 1:m)
{
  for(ind in 1:nfc)
  {
    individuo <- individuo + 1
    indrecm   <- c(rep(0,22))
    indrecp   <- indrecm
    gmmae     <- F0m[cF1[cross,1],] ## gameta materno da mãe
    gmpai     <- F0p[cF1[cross,1],] ## gameta paterno da mãe
    gmat      <- cbind(gmmae,gmpai)
    gmatr     <- cbind(gmpai,gmmae)
    gpmae     <- F0m[cF1[cross,2],] ## gameta materno do pai
    gppai     <- F0p[cF1[cross,2],] ## gameta paterno do pai
    gpat      <- cbind(gpmae,gppai)
    gpatr     <- cbind(gppai,gpmae)
    np1m     <- rpois(1,1)
    np2m     <- rpois(1,0.7)
    np1p     <- rpois(1,1)
    np2p     <- rpois(1,0.7)
    cr1m     <- 0
    cr2m     <- 0
    cr1p     <- 0
    cr2p     <- 0
    if(np1m>0)
    {
      pp1m <- runif(np1m)*100
    }
    if(np2m>0)
    {
      pp2m <- runif(np2m)*70
    }
    if(np1p>0)
    {
      pp1p <- runif(np1p)*100
    }
    if(np2p>0)
    {
      pp2p <- runif(np2p)*70
    }
    while(cr1m<np1m)
    {
      cr1m <- cr1m + 1
      for(i in 1:13)
      {
        if(pm[i] > pp1m[cr1m])
        {
          indrecm[i] <- indrecm[i] + 1
        }
      }
    }
    while(cr2m<np2m)
    {
      cr2m <- cr2m + 1
    }
  }
}

```

```

for(i in 14:22)
{
  if(pm[i] > pp2m[cr2m])
  {
    indrecm[i] <- indrecm[i] + 1
  }
}
while(cr1p<np1p)
{
  cr1p <- cr1p + 1
  for(i in 1:13)
  {
    if(pm[i] > pp1p[cr1p])
    {
      indrecp[i] <- indrecp[i] + 1
    }
  }
}
while(cr2p<np2p)
{
  cr2p <- cr2p + 1
  for(i in 14:22)
  {
    if(pm[i] > pp2p[cr2p])
    {
      indrecp[i] <- indrecp[i] + 1
    }
  }
}
indrecm <- indrecm %% 2
indrecp <- indrecp %% 2
UMm <- matrix(0, 22, 2)
UMp <- matrix(0, 22, 2)
UMc <- matrix(0, 22, 2)
for(i in 1:22)
{
  if (indrecm[i] == 0)
  {
    gmat[i,] <- gmat[i,]
    UMm[i,] <- c(1,0)
  }
  else
  {
    gmat[i,] <- gmatr[i,]
    UMm[i,] <- c(0,1)
  }
  if(indrecp[i] == 0)
  {
    gpat[i,] <- gpat[i,]
    UMp[i,] <- c(1,0)
  }
  else
  {
    gpat[i,] <- gpatr[i,]
    UMp[i,] <- c(0,1)
  }
}

```

```

    }
  }
  aux1      <- sample(1:2, 1)
  aux2      <- sample(1:2, 1)
  UMc[ , 1] <- UMc[ , aux1]
  UMc[ , 2] <- UMc[ , aux2]
  F1m[individuo,] <- t(gmat[ ,aux1])
  F1p[individuo,] <- t(gpat[ ,aux2])
  UMF1      <- rbind(UMc[1:3, ], UMc[5:9, ], UMc[11:16, ],
                    UMc[18:22, ])
  write(UMF1, file= "UmarcasF1.txt", 19, append=TRUE)
  UqtlF1    <- cbind(UMc[4,1],UMc[4 ,2],UMc[10,1],UMc[10,2],
                    UMc[17 ,1], UMc[17 ,2])
  write(UqtlF1, file = "UqtlF1.txt", 6, append=TRUE)
}
}
marcm1 <- rbind(F0m,F1m)
marcp1 <- rbind(F0p,F1p)
##### Geração F2 #####
##### Gerando estrutura de cruzamentos F2
cF2    <- matrix(0,m,2)
cF2[,1] <- sample((m+1):(nF1+m),m,replace=TRUE)
cF2[,2] <- sample((m+1):(nF1+m),m,replace=TRUE)
F2m    <- matrix(0,(m*nfc),22)
F2p    <- matrix(0,(m*nfc),22)
indrecm <- c(rep(0,22))
indrecp <- indrecm
gmmae  <- 0*F0m[1,]
gmpai  <- 0*F0p[1,]
gmat   <- cbind(gmmae,gmpai)
gmatr  <- gmat
gpmae  <- 0*F0m[1,]
gppai  <- 0*F0p[1,]
gpat   <- cbind(gpmae,gppai)
gpatr  <- gpat
indivduo <- 0
for(cross in 1:m)
{
  for(ind in 1:nfc)
  {
    individuo <- individuo + 1
    indrecm   <- c(rep(0,22))
    indrecp   <- indrecm
    gmmae    <- marcm1[cF2[cross,1],]
    gmpai    <- marcp1[cF2[cross,1],]
    gmat     <- cbind(gmmae,gmpai)
    gmatr    <- cbind(gmpai,gmmae)
    gpmae    <- marcm1[cF2[cross,2],]
    gppai    <- marcp1[cF2[cross,2],]
    gpat     <- cbind(gpmae,gppai)
    gpatr    <- cbind(gppai,gpmae)
    np1m    <- rpois(1,1)
    np2m    <- rpois(1,0.7)
    np1p    <- rpois(1,1)
    np2p    <- rpois(1,0.7)
    cr1m    <- 0
  }
}

```

```

cr2m <- 0
cr1p <- 0
cr2p <- 0
if(np1m>0)
{
  pp1m <- runif(np1m)*100
}
if(np2m>0)
{
  pp2m <- runif(np2m)*70
}
if(np1p>0)
{
  pp1p <- runif(np1p)*100
}
if(np2p>0)
{
  pp2p <- runif(np2p)*70
}
while(cr1m<np1m)
{
  cr1m <- cr1m + 1
  for(i in 1:13)
  {
    if(pm[i] > pp1m[cr1m])
    {
      indrecm[i] <- indrecm[i] + 1
    }
  }
}
while(cr2m<np2m)
{
  cr2m <- cr2m + 1
  for(i in 14:22)
  {
    if(pm[i] > pp2m[cr2m])
    {
      indrecm[i] <- indrecm[i] + 1
    }
  }
}
while(cr1p<np1p)
{
  cr1p <- cr1p + 1
  for(i in 1:13)
  {
    if(pm[i] > pp1p[cr1p])
    {
      indrecp[i] <- indrecp[i] + 1
    }
  }
}
while(cr2p<np2p)
{
  cr2p <- cr2p + 1
  for(i in 14:22)

```

```

    {
      if(pm[i] > pp2p[cr2p])
      {
        indrecp[i] <- indrecp[i] + 1
      }
    }
  }
  indrecm <- indrecm %% 2
  indrecp <- indrecp %% 2
  Umm      <- matrix(0, 22, 2)
  Ump      <- matrix(0, 22, 2)
  Umc      <- matrix(0, 22, 2)
  for(i in 1:22)
  {
    if (indrecm[i] == 0)
    {
      gmat[i,]      <- gmat[i,]
      Umm[i,1]      <- 1
      Umm[i,2]      <- 0
    }
    else
    {
      gmat[i,] <- gmatr[i,]
      Umm[i,1] <- 0
      Umm[i,2] <- 1
    }
    if (indrecp[i] == 0)
    {
      gpat[i,]      <- gpat[i,]
      Ump[i,1]      <- 1
      Ump[i,2]      <- 0
    }
    else
    {
      gpat[i,]      <- gpatr[i,]
      Ump[i,1]      <- 0
      Ump[i,2]      <- 1
    }
  }
  aux1      <- sample(1:2, 1)
  aux2      <- sample(1:2, 1)
  Umc[ , 1] <- Umm[ , aux1]
  Umc[ , 2] <- Ump[ , aux2]
  F2m[individuo,] <- t(gmat[ ,aux1])
  F2p[individuo,] <- t(gpat[ ,aux2])
  UMF2      <- rbind(UMc[1:3, ],UMc[5:9, ],UMc[11:16, ],
                    Umc[18:22, ])
  Uqt1F2    <- cbind(UMc[4 ,1],UMc[4 ,2], Umc[10 ,1],
                    Umc[10 ,2], Umc[17 ,1], Umc[17 ,2])
  write(UMF2, file= "UmarcasF2.txt", 19,append=TRUE)
  write(Uqt1F2, file = "Uqt1F2.txt", 6, append=TRUE)
}
}
marcm      <- rbind(F0m,F1m,F2m)
marcp      <- rbind(F0p,F1p,F2p)
aQTLm      <- cbind(marcm[,4], marcm[ ,10], marcm[ ,17])

```

```

aQTLp      <- cbind(marcp[,4], marcp[,10], marcp[,17])
marcamaterno <- cbind(marcm[,1:3],marcm[,5:9],marcm[,11:16],marcm[,18:22])
marcapaterno <- cbind(marcp[,1:3],marcp[,5:9],marcp[,11:16],marcp[,18:22])
colunasM    <- length(marcm[1,])-length(aQTLm[1,])
nQTL        <- length(aQTLm[1,])
MarcasTab   <- matrix(0,(2*n),colunasM)
for(i in 1:n)
{
  MarcasTab[(2*(i-1)+1),] <- marcamaterno[i,]
  MarcasTab[(2*(i-1)+2),] <- marcapaterno[i,]
}
write(t(MarcasTab), file= "marcas.txt", colunasM)
write(t(marcamaterno), file= "marcasm.txt", colunasM)
write(t(marcapaterno), file= "marcasp.txt", colunasM)
write(t(cF1), file= "cruzamentoF1.txt", 2)
write(t(cF2), file= "cruzamentoF2.txt", 2)
## definindo efeitos aditivos e de dominância dos QTLs
va1 <- 0.15
vd1 <- 0.10
va2 <- 0.10
vd2 <- 0.05
va3 <- 0.15
vd3 <- 0.10
a1 <- rnorm((2*m),0,sqrt(va1))
a2 <- rnorm((2*m),0,sqrt(va2))
a3 <- rnorm((2*m),0,sqrt(va3))
d1 <- matrix(0, 2*m, 2*m)
d2 <- d1
d3 <- d1
for (i in 1:(2*m))
{
  for (j in i:(2*m))
  {
    d1[i,j] <- rnorm(1,0,sqrt(vd1))
    d1[j,i] <- d1[i,j]
    d2[i,j] <- rnorm(1,0,sqrt(vd2))
    d2[j,i] <- d2[i,j]
    d3[i,j] <- rnorm(1,0,sqrt(vd3))
    d3[j,i] <- d3[i,j]
  }
}
a <- cbind(a1,a2,a3)
d <- cbind(d1,d2,d3)
write(a, file= "efaditivos.txt", length(a[1,]))
write(d, file= "efdominantes.txt",length(d[1,]))
maes <- c(c(rep(0,m)), kronecker(cF1[,1],c(rep(1,nfc))),
kronecker(cF2[,1],c(rep(1,nfc))))
pais <- c(c(rep(0,m)), kronecker(cF1[,2],c(rep(1,nfc))),
kronecker(cF2[,2],c(rep(1,nfc))))
#### valores paramétricos genotípicos e fenotípicos
ve <- 1
e <- sqrt(ve) * rnorm(n)
yg <- matrix(0,n,1)
for (ind in 1:n)
{
  yg[ind] <- a1[aQTLm[ind,1]] + a1[aQTLp[ind,1]] +

```

```

        a2[aQTLm[ind,2]] + a2[aQTLp[ind,2]] +
        a3[aQTLm[ind,3]] + a3[aQTLp[ind,3]] +
        d1[aQTLm[ind,1], aQTLp[ind,1]] +
        d2[aQTLm[ind,2], aQTLp[ind,2]] +
        d3[aQTLm[ind,3], aQTLp[ind,3]]
    }
y <- round((yg + e),4)
dados <- cbind(maes,pais,yg,y)
write(t(dados), file= "dadospgf.txt", 4)

```

Programa 2: Rotina R para definir os valores iniciais na primeira iteração.

```

##### definir os valores iniciais #####
library(MASS)
source("funcoesU.R")
Nef <- 50000
burn <- 1000
salto <- 10
maxiter <- burn + Nef*salto
pm<-c(0,10,20,30,40,50,60,70,80,90,100,100,110,120,130,140,150,160,170)
TC <- 1.7 ##### tamanho do cromossomo
delta <- .15
#### dados do pedigree #####
m <- 5
nfc <- 6
nF1 <- m*nfc
nF2 <- m*nfc
nF <- nF1+nF2
n <- m+nF
na <- 2*m
nd <- m*(2*m+1)
X <- matrix(1,n,1)
xtx <- ginv(t(X)%*%X)
xx <- xtx%*%t(X)
dados <- read.table("dadospgf.txt")
attach(dados)
y <- dados[,4]
maes <- dados[,1]
pais <- dados[,2]
marcasm <- read.table("marcasm.txt")
marcasp <- read.table("marcasp.txt")
#### hiperparâmetros das priors qui-quadrado
he <- 8
Se <- 10
ha <- 25
Sa <- 3
hd <- 25
Sd <- 3
##### Valores iniciais #####
ve <- 1
va <- c(0.15,0.10, 0.15, 0.15,0.10, 0.15)
vd <- c(0.1, 0.05, 0.10, 0.1, 0.05, 0.10)

```

```

b      <- 0
lam    <- c(0.125,0.372,1.643,0.8,0.9,1.4)
la     <- lam
Lc     <- 0
Lmax   <- 6
L      <- 6
mi     <- 2      ##### média da poisson, priori de L
Lmu    <- log(mi)
vUq    <- matrix(c(1,1,0,0,1,0,1,0), 4,2)
U      <- matrix(sample(0:1,nF*2*Lmax,replace=TRUE),(n-m),(2*Lmax))
Zm     <- matrix(0,n,(2*m))
Zp     <- matrix(0,n,(2*m))
W      <- matrix(0, n, m*(2*m+1))
a      <- matrix(rnorm(Lmax*2*m ,0,sqrt(0.14)), (2*m), Lmax)
d      <- matrix(rnorm(Lmax*nd ,0,sqrt(0.10)), nd, Lmax)
md     <- matrix(0, 2*m, 2*m)
mdc    <- md
##### variáveis auxiliares #####
auxL1  <- 0
auxL2  <- 0
A      <- 0
f      <- matrix(runif(m),m,1)
f      <- f*0
cross  <- 1
aux    <- 1
pad    <- 0
acand  <- a
dcand  <- d
vacand <- va
vdcand <- vd
dc     <- d
lamc   <- lam
pU     <- c(rep(0,Lmax))
pUc    <- pU
fp     <- 0
fpc    <- 0
pr     <- matrix(0,nF,Lmax)
prc    <- pr
lpr    <- 0
lprc   <- lpr
Zmp    <- matrix(0,n,na*Lmax)
Wq     <- matrix(0,n,nd*Lmax)
av     <- 0
dv     <- 0
iterqtl <- c(rep(0,Lmax))
### Valores iniciais para as matrizes Zp, Zm e W #####
for (ind in 1:m)
{
  Zm[ind, (2*ind - 1)] <- -1
  Zp[ind, (2*ind )]   <- -1
}
for (q in 1:3)
{
  for ( ind in (m+1):n)
  {
    Zobj <- ZdadoU(ind,U,q,Zm,Zp,maes,pais)
  }
}

```

```

Zm      <- Zobj$Zm
Zp      <- Zobj$Zp
}
Z        <- Zm + Zp
W        <- matrix(0, n, m*(2*m+1))
for (ind in 1: n)
{
  Wobj   <- WdadoZ(ind,Z,W)
  W      <- Wobj$W
}
Zmp[, (na*(q-1)+1):(na*q)] <- Z
Wq[, ((nd*(q-1)+1):(nd*q))] <- W
av[(na*(q-1)+1):(na*q)]   <- a[ ,q]
dv[(nd*(q-1)+1):(nd*q)]   <- d[ ,q]
}
for (q in 4:Lmax)
{
  Zmp[, (na*(q-1)+1):(na*q)] <- Zmp[, (na*(q-1-3)+1):(na*(q-3))]
  Wq[, ((nd*(q-1)+1):(nd*q))] <- Wq[, ((nd*(q-1-3)+1):(nd*(q-3)))]
  av[(na*(q-1)+1):(na*q)]   <- av[(na*(q-1-3)+1):(na*(q-3))]
  dv[(nd*(q-1)+1):(nd*q)]   <- dv[(nd*(q-1-3)+1):(nd*(q-3))]
}
##### Alterar o número de QTLs para 3 #####
L      <- 3
for (q in 4:Lmax)
{
  lam[q] <- 0
  a[ ,q] <- a[ ,1]*0
  d[ ,q] <- d[ ,1]*0
  va[q]  <- 0
  vd[q]  <- 0
  av[(na*(q-1)+1):(na*q)] <- a[ ,q]*0
  dv[(nd*(q-1)+1):(nd*q)] <- d[ ,q]*0
}
##### Definir todos os valores em uma primeira iteração #####
for (q in 1:3 )
{
  if (lam[q] > 0 )
  {
    la <- lam[q]
    objMatrizesED <- MatrizesED(lam,Lmax,q,la,pm,U)
    Umarca <- objMatrizesED$Umarca
    r1 <- objMatrizesED$r1
    r2 <- objMatrizesED$r2
    pr1 <- matrix(0,nF,4)
    auxvU <- 0
    probr1 <- matrix(0, nF, 8)
    probr2 <- matrix(0, nF, 8)
    Zmc <- Zm
    Zpc <- Zp
    Zc <- Z
    Zcm <- matrix(0,4,2*m)
    Zcp <- matrix(0,4,2*m)
    Wac <- W[1:4, ]
    Wc <- W
    Zmpca <- Zmp
  }
}

```

```

Wca          <- Wq
Zmpcand      <- Zmp
Wqcand       <- Wq
for (individuo in (m+1): n)
{
  ind        <- individuo - m
  veross     <- c(rep(0,4))
  for (t in 1 : 4) ##### 4 opcoes para o vetor U
  {
    ### fazer o calculo de cada linha das matrizes Z e W
    Zlinhaobj <- linhaZ(individuo,vUq,t,q,Zm,Zp,maes,pais)
    Zmc       <- Zlinhaobj$Zmc
    Zpc       <- Zlinhaobj$Zpc
    Zc[individuo, ] <- Zmc[individuo, ] + Zpc[individuo, ]
    Wc[individuo, ] <- W[individuo, ]*0
    Wobj      <- WdadoZ(individuo,Zc,Wc)
    Wc        <- Wobj$W
    Zmpca[individuo,(na*(q-1)+1):(na*q)] <- Zc[individuo, ]
    Wca[individuo,((nd*(q-1)+1):(nd*q))] <- Wc[individuo, ]
    ##### verossimilhança para cada linha
    S1ca      <- Zmpca[individuo,]*%av + Wca[individuo,]*%dv
    R1ca      <- y[individuo] - b - S1ca
    fyc       <- exp(- (R1ca^2)/(2*ve ))
    veross[t] <- fyc
    Zcm[t, ]  <- Zmc[individuo, ]
    Zcp[t, ]  <- Zpc[individuo, ]
    Wac[t, ]  <- Wc[individuo, ]
    objPrGenQTL <- PrGenQTL(r1,r2,vUq,t,Umarca,pr1,probr1, probr2,ind)
    pr1       <- objPrGenQTL$pr1
    probr1    <- objPrGenQTL$probr1
    probr2    <- objPrGenQTL$probr2
  }
  ##### fecha for t
  ##### sortear o u candidato para o individuo #####
  pu         <- pr1[ind,]*veross/sum(pr1[ind,]*veross)
  uv         <- c(1,2,3,4)
  auxU       <- sample(uv,1,prob=pu)
  Zm[individuo, ] <- Zcm[auxU, ]
  Zp[individuo, ] <- Zcp[auxU, ]
  W[individuo, ] <- Wac[auxU, ]
  Zmpcand[individuo,(na*(q-1)+1):(na*q)] <- Zm[individuo,]+Zp[individuo,]
  Wqcand[individuo,((nd*(q-1)+1):(nd*q))] <- W[individuo,]
  ### indicadores de segregação do QTL para os descendentes.
  U[ind, (2*q-1):(2*q)] <- vUq[auxU, ]
  auxvU[ind] <- auxU
  prc[ind ,q] <- pr1[ind,auxU]/sum(pr1[ind,])
}
##### fecha for individuo
lpr[q]      <- sum(log(prc[,q]))
##### Calcular a Verossimilhança no valor atual
S1c         <- Zmpcand%*av + Wqcand%*dv
Rc          <- y - X%*b - S1c
lfyc        <- - (t(Rc)%*Rc)/(2*ve )
##### Verossimilhança no valor candidato
S1          <- Zmp%*av + Wq%*dv
R           <- y - X%*b - S1
lfy         <- - (t(R)%*R)/(2*ve )
lAf         <- (lfyc -lfy)

```

```

ifelse(lAf>0, A <- 1, A <- exp(lAf))
if (A > runif(1))
{
  lam[q] <- la
  Zmp[ , (na*(q-1)+1):(na*q)] <- Zmpcand[ , (na*(q-1)+1):(na*q)]
  Wq[ , ((nd*(q-1)+1):(nd*q))] <- Wqcand[ , ((nd*(q-1)+1):(nd*q))]
}
lamc[q] <- lam[q] ##### vetor lambda diferente de zero
} ##### fechando if (lam[q] > 0 )
} ##### (fechando for q)
for (q in 4:Lmax)
{
  Zmp[, (na*(q-1)+1):(na*q)] <- Zmp[, (na*(q-1-3)+1):(na*(q-3))]
  Wq[, ((nd*(q-1)+1):(nd*q))] <- Wq[, ((nd*(q-1-3)+1):(nd*(q-3)))]
  lpr[q] <- lpr[(q-3)]
}
lprc <- lpr
### contadores
contla <- c(rep(0,Lmax))
contva <- c(rep(0,Lmax))
contvd <- c(rep(0,Lmax))
contad <- matrix(0, Lmax, m)
contdom <- matrix(0, Lmax, m*m)
iterqtl <- c(rep(0,Lmax))

```

Programa 3: Rotina R para definir as funções invocadas no programa de análise.

```

##### Inicia as funcoes #####
#### matrizes de herança alélica #####
ZdadoU <- function(ind,U,q,Zm,Zp,maes,pais)
{
  if ( U[(ind-m),(2*q-1)] == 1 ) ##### alelo paterno da mãe
  {
    Zm[ind, ] <- Zm[maes[ind], ]
  }
  else
  {
    Zm[ind, ] <- Zp[maes[ind], ]
  }
  if (U[(ind-m),(2*q)] == 1 ) ##### alelo paterno do pai
  {
    Zp[ind, ] <- Zm[pais[ind], ]
  }
  else
  {
    Zp[ind, ] <- Zp[pais[ind], ]
  }
  list(Zm=Zm,Zp=Zp)
}
#### matrizes de herança alélica por linha #####

```

```

linhaZ <- function(individuo,vUq,t,q,Zm,Zp,maes,pais)
{
  if ( vUq[t,1] == 1 )   ### alelo materno da mãe
  {
    Zmc[individuo, ] <- Zm[maes[individuo], ]
  } else
  {
    Zmc[individuo, ] <- Zp[maes[individuo], ]
  }
  if (vUq[t,2] == 1 )   ### alelo materno do pai
  {
    Zpc[individuo, ] <- Zm[pais[individuo], ]
  } else
  {
    Zpc[individuo, ] <- Zp[pais[individuo], ]
  }
  list(Zmc=Zmc,Zpc=Zpc)
}
##### matriz de delineamento dos efeitos de dominância #####
WdadoZ <- function(ind,Z,W)
{
  if ( Z[ind, 1] == 2 )
  {
    W[ind, 1] <- 1
  }
  else
  {
    if ( Z[ind, 1] == 1 )
    {
      for (j in 2:(2*m))
      {
        if ( Z[ind, j] == 1 )
        {
          W[ind, j] <- 1
        }
      }
    }
  }
  auxW1 <- 2*m
  for (j in 2:(2*m))
  {
    if ( Z[ind, j] == 2 )
    {
      W[ind, (auxW1 +1)] <- 1
    }
    auxW1 <- auxW1 + 2*m - (j-1)
  }
  auxW2 <- 2*m
  for (i in 2:(2*m-1))
  {
    if ( Z[ind, i] == 1 )
    {
      for (j in (i+1): (2*m))
      {
        if ( Z[ind, j] == 1 )
        {

```

```

        W[ind, auxW2 + j-(i-1)] <- 1
      }
    }
  }
  auxW2 <- auxW2 + 2*m - (i-1)
}
list(W=W)
}
##### distâncias à esquerda e à direita (função de Haldane )
MatrizesED <- function(lam,Lmax,q,la,pm,U)
{
  marcaesq <- matrix(0,n,2)
  marcadir <- matrix(0,n,2)
  nF <- length(marcaesq[,1]) - m
  Umarca <- matrix(0,(2*nF),2)
  ifelse( la < 1, j <- floor(10*la) + 1 ,j <- floor(10*la) + 2)
  if(la == 1.7 )
  {
    j <- 18
  }
  if (la == 1)
  {
    t1 <- c(10,12)
    j <- sample(t1,1)
  }
  marcaesq[,1] <- marcasq[, j]
  marcaesq[,2] <- marcasq[, j]
  marcadir[,1] <- marcasq[, (j+1)]
  marcadir[,2] <- marcasq[, (j+1)]
  esq <- la - pm[j]/100
  dir <- pm[j+1]/100 - la
  auxe1 <- 0
  auxe2 <- 0
  for (j in 1 :Lmax)
  {
    if ( j!=q)
    {
      e1 <- la - lam[j]
      e2 <- lam[j] - la
      if ( (e1>0) && (e1 < esq))
      {
        esq <- e1
        auxe1 <- 1
      }
      if ( (e2>0) && (e2 < dir))
      {
        dir <- e2
        auxe2 <- 1
      }
    }
  }
  r1 <- (1-exp(-2*esq))/2
  r2 <- (1-exp(-2*dir))/2
  for (ind in 1:nF)
  {
    if (marcaesq[maes[(ind+m)], 1] == marcaesq[maes[(ind+m)], 2] )

```

```

    {
      Umarca[(2*ind-1),1] <- sample(0:1,1)
    }
  else
  {
    if (marcaesq[(ind+m), 1] == marcaesq[maes[(ind+m)], 1] )
    {
      Umarca[(2*ind-1),1] <- 1
    }
  }
  if (marcaesq[pais[(ind+m)], 1] == marcaesq[pais[(ind+m)], 2] )
  {
    Umarca[(2*ind),1] <- sample(0:1,1)
  }
  else
  {
    if (marcaesq[(ind+m), 2] == marcaesq[pais[(ind+m)], 1] )
    {
      Umarca[(2*ind),1] <- 1
    }
  }
  if (marcadir[maes[(ind+m)], 1] == marcadir[maes[(ind+m)], 2] )
  {
    Umarca[(2*ind-1),2] <- sample(0:1,1)
  }
  else
  {
    if (marcadir[(ind+m), 1] == marcadir[maes[(ind+m)], 1] )
    {
      Umarca[(2*ind-1),2] <- 1
    }
  }
  if (marcadir[pais[(ind+m)], 1] == marcadir[pais[(ind+m)], 2] )
  {
    Umarca[(2*ind),2] <- sample(0:1,1)
  }
  else
  {
    if (marcadir[(ind+m), 2] == marcadir[pais[(ind+m)], 1] )
    {
      Umarca[(2*ind),1] <- 1
    }
  }
}
if (auxe1==1)
{
  Umarca[(2*ind-1):(2*ind),1] <- U[ind, (2*q-1):(2*q)]
}
if (auxe2==1)
{
  Umarca[(2*ind-1):(2*ind),2] <- U[ind, (2*q-1):(2*q)]
}
} ### for ind
list(Umarca=Umarca,r1=r1,r2=r2)
}
#### probabilidades de recombinação entre QTL e marcadores
PrGenQTL <- function(r1,r2,vUq,t,Umarca,pr1,probr1,probr2,ind)

```

```

{
  ifelse(vUq[t, 1]== Umarca[(2*ind-1),1],probr1[ind,(2*t-1)]<-(1-r1),
    probr1[ind,(2*t-1)] <- r1)
  ifelse(vUq[t, 2]== Umarca[(2*ind),1], probr1[ind,(2*t)] <-(1-r1),
    probr1[ind,(2*t)] <- r1)
  ifelse(vUq[t, 1]== Umarca[(2*ind-1),2],probr2[ind,(2*t-1)]<-(1-r2),
    probr2[ind,(2*t-1)] <- r2)
  ifelse(vUq[t, 2]== Umarca[(2*ind),2], probr2[ind,(2*t)] <-(1-r2),
    probr2[ind,(2*t)] <- r2)
  pr1[ind,1] <-probr1[ind,1]*probr1[ind,2]* probr2[ind , 1]*probr2[ind , 2]
  pr1[ind,2] <-probr1[ind,3]*probr1[ind,4]* probr2[ind , 3]*probr2[ind , 4]
  pr1[ind,3] <-probr1[ind,5]*probr1[ind,6]* probr2[ind , 5]*probr2[ind , 6]
  pr1[ind,4] <-probr1[ind,7]*probr1[ind,8]* probr2[ind , 7]*probr2[ind , 8]
  list(pr1=pr1,probr1=probr1,probr2=probr2)
}
##### atualizar efeitos aditivos (Metropolis Hastings)
akdadofk <- function(q,Zmp,Wq,av,dv,Lmax,lam,f,a,va,X,b,contad,acand)
{
  S1 <-Zmp%*av + Wq%*dv
  R <- y - X%*b - S1
  lfyant <- - (t(R)%*R)/(2*ve )
  for (k in 1: m)
  {
    delta1 <- runif(1,-0.8,0.8)
    delta2 <- runif(1,-0.8,0.8)
    cov <- matrix(f[k,1], 2, 2)
    cov[1,1] <- 1
    cov[2,2] <- 1
    V_ak <- va[q]*cov
    ac <- a
    avc <-av
    ## ak: vetor com 2 efeitos alélicos do kesimo fundador
    ak <- a[((2*k-1):(2*k)), q]
    ## c: vetor candidato com 2 efeitos alélicos do kesimo fundador
    c <- ak + c(delta1, (f[k,1]*delta1 + (1- f[k,1])*delta2))
    ## ac: vetor candidato com 2m efeitos alélicos
    ac[((2*k-1):(2*k)), q] <- c
    avc[(na*(q-1)+1):(na*q)] <- ac[ ,q]
    S1c <-Zmp%*avc + Wq%*dv
    Rc <- y - X%*b - S1c
    lfyacand <- - (t(Rc)%*Rc)/(2*ve )
    lpant <- -(t(ak)%*ginv(V_ak)%*ak)/2
    lpcand <- -(t(c)%*ginv(V_ak)%*c)/2
    lAf <- (lfyacand+lpcand-lfyant-lpant)
    ifelse(lAf>0, A <- 1, A <- exp(lAf))
    if (A > runif(1))
    {
      lfyant <- lfyacand
      a <- ac
      av <- avc
      contad[q, k] <- contad[q, k] +1
    }
  }
  ## (fechando for k in 1:m)
  acand[ ,q] <- a[ ,q]
  list(a=a,av=av,acand=acand,lfyant=lfyant, contad=contad)
}

```

```

##### Atualizar efeitos de dominancia (Metropolis Hastings)
dkdadovd <- function(q,Zmp,Wq,av,dv,Lmax,lam,f,d,vd,X,b,contdom,dcand,lfyant)
{
  md      <- matrix(0, 2*m, 2*m)
  mdc     <- md
  cnd     <- 0
  for (k in 1:m)
  {
    for (kk in k: m)
    {
      cnd      <- cnd + 1
      del1     <- runif(1,-0.8,0.8)
      del2     <- runif(1,-0.8,0.8)
      del3     <- runif(1,-0.8,0.8)
      del4     <- runif(1,-0.8,0.8)
      auxd     <- 2*m
      md[1, ] <- d[1:(2*m), q]
      for (j in 2:(2*m))
      {
        b1      <- auxd + 2*m - (j-1)
        md[j, j:(2*m)] <- d[(auxd+1): b1, q]
        auxd    <- b1
      }
      mdc <- md
      dc <- d
      dvc <- dv
      if ( k == kk)
      {
        cov    <- diag(3)
        Vdk    <- vd[q] * cov
        dk     <- c(md[(2*k-1), (2*k-1)], md[(2*k-1), (2*k)], md[(2*k), (2*k)])
        dkc    <- dk + c(del1, f[k]*del1+(1-f[k])*del2, f[k]*del1+(1-f[k])*del4)
        mdc[(2*k-1), (2*k-1)] <- dkc[1]
        mdc[(2*k-1), (2*k)] <- dkc[2]
        mdc[(2*k), (2*k)] <- dkc[3]
      }
      else
      {
        cov <- diag(4)
        Vdk <- vd[q] * cov
        dk <- c(md[(2*k-1), (2*kk-1)], md[(2*k-1), (2*kk)], md[(2*k), (2*kk-1)], md[(2*k), (2*kk)])
        dkc <- dk + c(del1, f[k]*del1 + (1-f[k])*del2, f[kk]*del1 + (1-f[kk])*del3, f[k]*f[kk]*del1 + (1-f[k])*f[kk]*del2 + f[k]*(1-f[kk])*del3 + (1-f[k])*(1-f[kk])*del4 )
        mdc[(2*k-1), (2*kk-1)] <- dkc[1]
        mdc[(2*k-1), (2*kk)] <- dkc[2]
        mdc[(2*k), (2*kk-1)] <- dkc[3]
        mdc[(2*k), (2*kk)] <- dkc[4]
      }
      daux      <- mdc[1,1:(2*m)]
      for ( j in 2: (2*m))
      {
        daux      <- c(cbind(t(daux), t(mdc[j, j:(2*m)])) )
      }
      dc[ , q] <- daux
    }
  }
}

```



```

Wca          <- Wq
Zmpcand     <- Zmp
Wqcand      <- Wq
for (individuo in (m+1): n)
{
  ind          <- individuo - m
  veross      <- c(rep(0,4))
  for (t in 1 : 4) ##### 4 opcoes para o vetor U
  {
    ##### fazer o calculo de cada linha das matrizes Z e W
    Zlinhaobj <- linhaZ(individuo,vUq,t,q,Zm,Zp,maes,pais)
    Zmc      <- Zlinhaobj$Zmc
    Zpc      <- Zlinhaobj$Zpc
    Zc[individuo, ] <- Zmc[individuo, ] + Zpc[individuo, ]
    Wc[individuo, ] <- W[individuo, ]*0
    Wobj     <- WdadoZ(individuo,Zc,Wc)
    Wc       <- Wobj$W
    Zmpca[individuo,(na*(q-1)+1):(na*q)] <- Zc[individuo, ]
    Wca[individuo,((nd*(q-1)+1):(nd*q))] <- Wc[individuo, ]
    ##### verossimilhança para cada linha
    S1ca     <- Zmpca[individuo, ]*%av + Wca[individuo, ]*%dv
    R1ca     <- y[individuo] - b - S1ca
    fyc      <- exp(- (R1ca^2)/(2*ve ) )
    veross[t] <- fyc
    Zcm[t, ] <- Zmc[individuo, ]
    Zcp[t, ] <- Zpc[individuo, ]
    Wac[t, ] <- Wc[individuo, ]
    objPrGenQTL <- PrGenQTL(r1,r2,vUq,t,Umarca,pr1,probr1, probr2,ind)
    pr1       <- objPrGenQTL$pr1
    probr1    <- objPrGenQTL$probr1
    probr2    <- objPrGenQTL$probr2
  }
  ##### u candidato para o ind, com as respectivas probabilidades
  pu <- pr1[ind,]*veross/sum(pr1[ind,]*veross)
  uv <- c(1,2,3,4)
  auxU <- sample(uv,1,prob=pu)
  Zm[individuo, ] <- Zcm[auxU, ]
  Zp[individuo, ] <- Zcp[auxU, ]
  W[individuo, ] <- Wac[auxU, ]
  Zmpcand[individuo,(na*(q-1)+1):(na*q)] <- Zm[individuo,]+Zp[individuo,]
  Wqcand[individuo,((nd*(q-1)+1):(nd*q))] <- W[individuo,]
  ### indicadores de segregação do QTL para os descendentes.
  U[ind,(2*q-1):(2*q)] <- vUq[auxU, ]
  auxvU[ind] <- auxU
  prc[ind ,q] <- pr1[ind,auxU]/sum(pr1[ind,])
}
lprc[q] <- sum(log(prc[,q]))
S1c <- Zmpcand*%av+Wqcand*%dv ### logverossimilhança valor atual
Rc <- y - X*%b - S1c
lfyc <- - (t(Rc)*%Rc)/(2*ve )
S1 <- Zmp*%av + Wq*%dv
R <- y - X*%b - S1
lfy <- - (t(R)*%R)/(2*ve ) ### logverossimilhança valor candidato
lAf <- (lfyc-lfy)
ifelse(lAf>0, A <- 1, A <- exp(lAf))
if (A > runif(1))

```

```

{
  lam[q] <- la
  contla[q] <- contla[q] + 1
  lpr[q] <- lprc[q]
  pr[,q] <- prc[,q]
  S1 <- S1c
  R <- Rc
  lfy <- lfy
  Zmp[ ,(na*(q-1)+1):(na*q)] <- Zmpcand[ ,(na*(q-1)+1):(na*q)]
  Wq[ ,(nd*(q-1)+1):(nd*q)] <- Wqcand[ ,(nd*(q-1)+1):(nd*q)]
}
lamc[q] <- lam[q] ##### vetor lambda diferente de zero
##### Amostrador de Gibbs para a e ve #####
aq <- av[(na*(q-1)+1):(na*q)]
va[q] <- (t(aq)%*%aq + Sa)/rchisq(1,(na+ha))
##### Amostrador akj (Metropolis Hastings)
akobj <- akdadofk(q,Zmp,Wq,av,dv,Lmax,lam,f,a,va,X,b,contad,acand)
a <- akobj$a
acand <- akobj$acand
lfy <- akobj$lfyant
contad <- akobj$contad
av <- akobj$av
### variancia de dominância vdj (amostrador de Gibbs) #####
dq <- dv[(nd*(q-1)+1):(nd*q)]
vd[q] <- (t(dq)%*%dq + Sd)/rchisq(1,(nd+hd))
dkobj <- dkdadovd(q,Zmp,Wq,av,dv,Lmax,lam,f,d,vd,X,b,contdom,dcand,lfy)
d <- dkobj$d
dcand <- dkobj$dcand
lfy <- dkobj$lfyant
contdom <- dkobj$contdom
dv <- dkobj$dv
iterqtl[q] <- iterqtl[q] + 1
} ##### fechando if (lam[q] > 0 )
##### fechando for q
##### Amostrador de Gibbs para b e ve
S1 <- Zmp%*%av + Wq%*%dv
mb <- xx%*%(y - S1)
vb <- txt*ve
b <- rnorm(1, mb, vb)#
R <- y - X%*%b - S1
ve <- (t(R)%*%R+Se)/rchisq(1,(n+he))
R <- y - X%*%b - S1
lfy <- - (t(R)%*%R)/(2*ve )

##### PASSO REVERSIBLE JUMP #####
#####
if (L==1)
{
  pad <- 1
}
if ( L== Lmax)
{
  pad <- 0
}
if (L > 1 && L < Lmax)
{

```

```

pad <- runif(1)
}
pa <- 0.5
pd <- 0.5
##### escolher um QTL para ser deletado do modelo #####
if (pad < 0.5)
{
Lc <- L-1
auxL1 <- sample(1:Lmax, 1)
while (lam[auxL1] == 0)
{
auxL1 <- sample(1:Lmax, 1)
}
avc <- av
dvc <- dvc
avc[(na*(auxL1-1)+1):(na*auxL1)] <- avc[(na*(auxL1-1)+1):(na*auxL1)]*0
dvc[(nd*(auxL1-1)+1):(nd*auxL1)] <- dvc[(nd*(auxL1-1)+1):(nd*auxL1)]*0
S1c <- Zmp*%avc + Wq*%dvc
Rc <- y - X*%b - S1c
lfyc <- - (t(Rc)%*Rc)/(2*ve )
LL <- log(L)
lAf <- (lfyc + LL - lfy - Lmu )
ifelse(lAf>0, A <- 1, A <- exp(lAf))
if (A > runif(1))
{
L <- Lc
av <- avc
dv <- dvc
a[ ,auxL1] <- avc[(na*(auxL1-1)+1):(na*auxL1)]
d[ ,auxL1] <- dvc[(nd*(auxL1-1)+1):(nd*auxL1)]
lam[auxL1] <- 0
va[auxL1] <- 0
vd[auxL1] <- 0
}
}
### Nascimento de um QTL (escolher uma posição para adicionar um QTL) ###
if (pad >= 0.5)
{
Lc <- L+1
auxL2 <- sample(1:Lmax, 1)
while (lam[auxL2] > 0)
{
auxL2 <- sample(1:Lmax, 1)
}
la1 <- lamc[auxL2]
lac <- round(runif(1,la1-delta, la1+delta),4)
la <- min(abs(lac), 2*TC -(abs(lac)) )
objMatrizesED <- MatrizesED(lam,Lmax,q,la,pm,U)
Umarca <- objMatrizesED$Umarca
r1 <- objMatrizesED$r1
r2 <- objMatrizesED$r2
pr1 <- matrix(0,nF,4)
auxvU <- 0
probr1 <- matrix(0, nF, 8)
probr2 <- matrix(0, nF, 8)
Zmc <- Zm
}
}

```

```

Zpc          <- Zp
Zc           <- Z
Zcm          <- matrix(0,4,2*m)
Zcp          <- matrix(0,4,2*m)
Wac         <- W[1:4, ]
Wc          <- W
Zmpca       <- Zmp
Wca         <- Wq
Zmpcand     <- Zmp
Wqcand      <- Wq
for (individuo in (m+1): n)
{
  ind        <- individuo - m
  veross     <- c(rep(0,4))
  for (t in 1 : 4) ##### 4 opcoes para o vetor U
  {
    Zlinhaobj <- linhaZ(individuo,vUq,t,auxL2,Zm,Zp,maes,pais)
    Zmc       <- Zlinhaobj$Zmc
    Zpc       <- Zlinhaobj$Zpc
    Zc[individuo, ] <- Zmc[individuo, ] + Zpc[individuo, ]
    Wc[individuo, ] <- W[individuo, ]*0
    Wobj      <- WdadoZ(individuo,Zc,Wc)
    Wc        <- Wobj$W
    Zmpca[individuo,(na*(auxL2-1)+1):(na*auxL2)] <- Zc[individuo, ]
    Wca[individuo,((nd*(auxL2-1)+1):(nd*auxL2))] <- Wc[individuo, ]
    S1ca      <- Zmpca[individuo, ]*%av + Wca[individuo, ]*%dv
    R1ca      <- y[individuo] - b - S1ca
    fyc       <- exp(- (R1ca^2)/(2*ve ) )
    veross[t] <- fyc
    Zcm[t, ]  <- Zmc[individuo, ]
    Zcp[t, ]  <- Zpc[individuo, ]
    Wac[t, ]  <- Wc[individuo, ]
    objPrGenQTL<- PrGenQTL(r1,r2,vUq,t,Umarca,pr1,probr1, probr2,ind)
    pr1       <- objPrGenQTL$pr1
    probr1    <- objPrGenQTL$probr1
    probr2    <- objPrGenQTL$probr2
  }
  pu         <- pr1[ind,]*veross/sum(pr1[ind,]*veross)
  uv         <- c(1,2,3,4)
  auxU      <- sample(uv,1,prob=pu)
  Zm[individuo, ] <- Zcm[auxU, ]
  Zp[individuo, ] <- Zcp[auxU, ]
  W[individuo, ] <- Wac[auxU, ]
  Zmpcand[individuo,(na*(auxL2-1)+1):(na*auxL2)]<-Zm[individuo,]+
  Zp[individuo,]
  Wqcand[individuo,((nd*(auxL2-1)+1):(nd*auxL2))]<- W[individuo,]
  U[ind, (2*auxL2-1):(2*auxL2)] <- vUq[auxU, ]
  auxvU[ind] <- auxU
  prc[ind ,auxL2] <- pr1[ind,auxU]/sum(pr1[ind,])
}
lprc[auxL2] <- sum(log(prc[,auxL2]))
### gerar um valor candidato para os efeitos alelicos dos fundadores
ac          <- a
avc        <-av
for (k in 1: m)
{

```

```

delta1 <- runif(1,-0.8,0.8)
delta2 <- runif(1,-0.8,0.8)
cov <- matrix(f[k,1], 2, 2)
cov[1,1] <- 1
cov[2,2] <- 1
V_ak <- va[q]*cov
ak <- a[((2*k-1):(2*k)), q]
c <- ak + c(delta1, (f[k,1]*delta1 + (1- f[k,1])*delta2))
ac[((2*k-1):(2*k)), auxL2] <- c
avc[(na*(auxL2-1)+1):(na*auxL2)] <- ac[,auxL2]
}
##### gerar um valor candidato para os efeitos de dominancia dj
aux <- 2*m
md[1, ] <- d[1:(2*m), auxL2]
for (j in 2:(2*m))
{
  b1 <- aux + 2*m - (j-1)
  md[j, j:(2*m)] <- d[(aux+1): b1, auxL2]
  aux <- b1
}
mdc <- md
dc <- d
dvc <- dv
for (k in 1:m)
{
  for (kk in k: m)
  {
    del1 <- runif(1,-0.8,0.8)
    del2 <- runif(1,-0.8,0.8)
    del3 <- runif(1,-0.8,0.8)
    del4 <- runif(1,-0.8,0.8)
    if ( k == kk)
    {
      dk <- c(md[(2*k-1), (2*k-1)], md[(2*k-1), (2*k)], md[(2*k), (2*k)])
      dkc<- dk+c(del1, f[k]*del1+(1-f[k])*del2, f[k]*del1+(1-f[k])*del4)
      mdc[(2*k-1), (2*k-1)] <- dkc[1]
      mdc[(2*k-1), (2*k)] <- dkc[2]
      mdc[(2*k), (2*k)] <- dkc[3]
    }
    else
    {
      dk <- c(md[(2*k-1), (2*kk-1)], md[(2*k-1), (2*kk)],
              md[(2*k), (2*kk-1)], md[(2*k), (2*kk)])
      dkc <- dk+c(del1, f[k]*del1+(1-f[k])*del2, f[kk]*del1+
                  (1-f[kk])*del3, f[k]*f[kk]*del1 + (1-f[k])*f[kk]*del2+
                  f[k]*(1-f[kk])*del3+(1-f[k])*(1-f[kk])*del4 )
      mdc[(2*k-1), (2*kk-1)] <- dkc[1]
      mdc[(2*k-1), (2*kk)] <- dkc[2]
      mdc[(2*k), (2*kk-1)] <- dkc[3]
      mdc[(2*k), (2*kk)] <- dkc[4]
    }
  }
  daux <- mdc[1,1:(2*m)]
  for ( j in 2: (2*m))
  {
    daux <- c(cbind(t(daux), t(mdc[j, j:(2*m)])) )
  }
}

```


Anexo B. Comparações entre as distribuições a priori e a posteriori para componentes de variância.

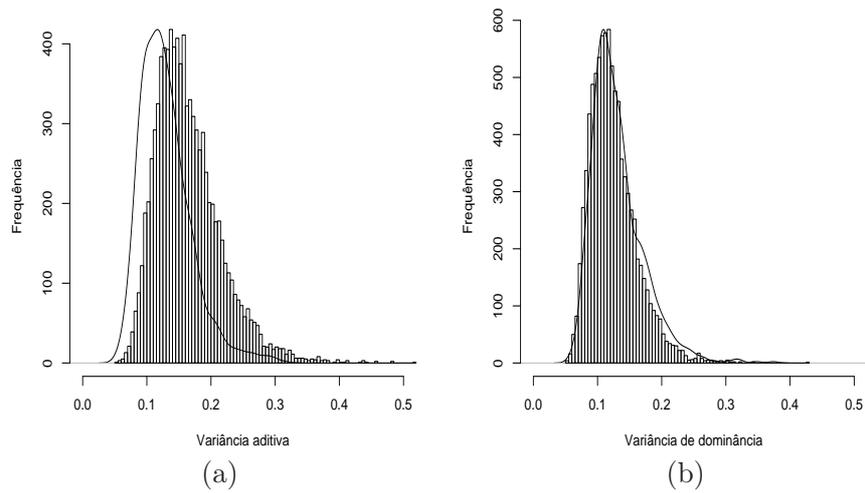


FIGURA 16: Distribuições a priori (densidade) e a posteriori (histograma) das variâncias aditiva e de dominância no intervalo de 30 a 40cM (*a* e *b*). (9.000 iterações, $nf=5$, $nfc=10$, $nc=1$, $nQTL=2$, $dM=10$).

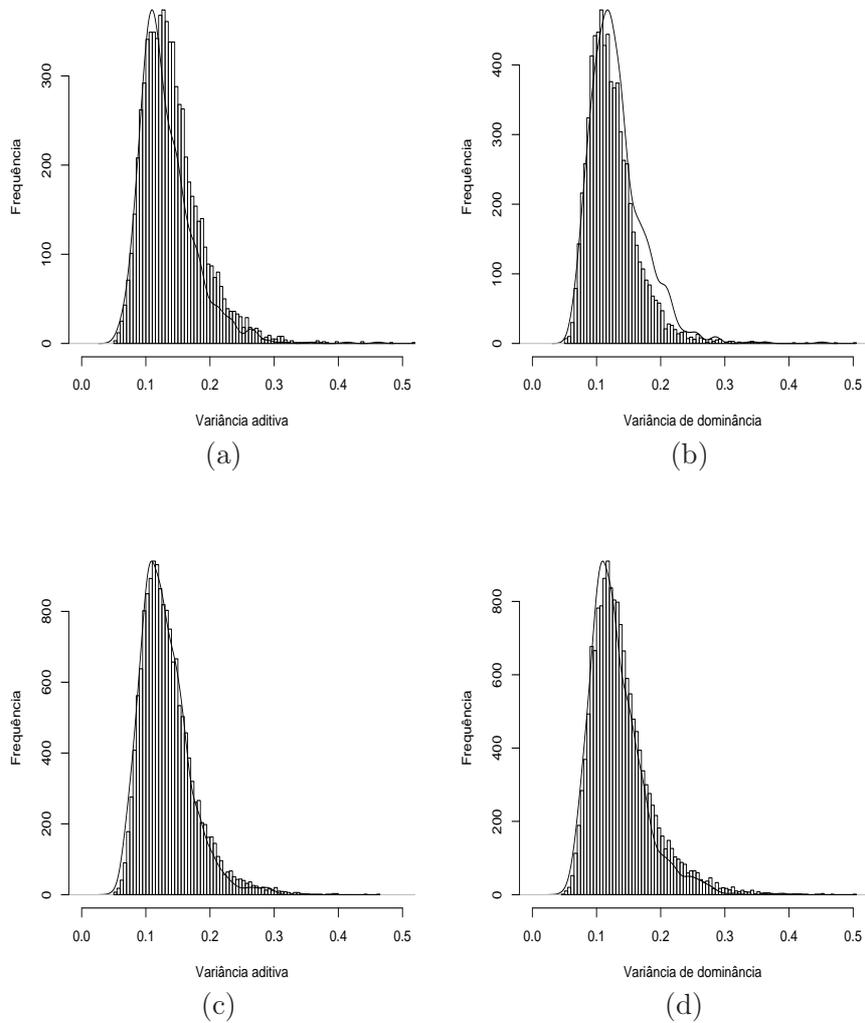


FIGURA 17: Distribuições a priori (densidade) e a posteriori (histograma) das variâncias aditiva e de dominância no intervalo de 30 a 40cM (*a* e *b*) e de 65 a 85cM (*c* e *d*). (24.000 iterações, $nf=5$, $nfc=10$, $nc=1$, $nQTL=2$, $dM=2$).

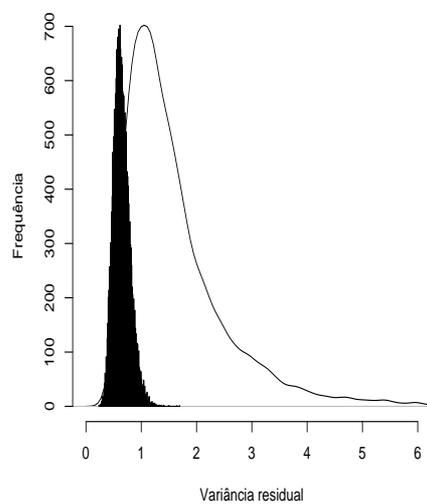


FIGURA 18: Distribuições a priori (densidade) e a posteriori (histograma) para a variância residual. (24.000 iterações, $nf=5$, $nfc=10$, $nc=1$, $nQTL=2$, $dM=2$).