

Universidade de São Paulo
Escola Superior de Agricultura "Luiz de Queiroz"

Mapeamento de QTLs utilizando as
abordagens Clássica e Bayesiana

Elisabeth Regina de Toledo

Dissertação apresentada para obtenção do título
de Mestre em Agronomia. Área de concentração:
Estatística e Experimentação Agronômica

Piracicaba

2006

Livros Grátis

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.

Elisabeth Regina de Toledo

Estatística

**Mapeamento de QTLs utilizando as
abordagens Clássica e Bayesiana**

Orientadora:

Profa. Dra. **ROSELI APARECIDA LEAN-
DRO**

Dissertação apresentada para obtenção do título
de Mestre em Agronomia. Área de concentração:
Estatística e Experimentação Agronômica

Piracicaba

2006

**Dados Internacionais de Catalogação na Publicação (CIP)
DIVISÃO DE BIBLIOTECA E DOCUMENTAÇÃO - ESALQ/USP**

Toledo, Elisabeth Regina de
Mapeamento de QTLs utilizando as abordagens Clássica e Bayesiana / Elisabeth
Regina de Toledo. - - Piracicaba, 2006.
99 p.

Dissertação (Mestrado) - - Escola Superior de Agricultura Luiz de Queiroz, 2006.

1. Estatística aplicada 2. Grãos 3. Mapeamento genético 4. Marcador molecular
5. Milho 6. Verossimilhança I. Título

CDD 633.15

“Permitida a cópia total ou parcial deste documento, desde que citada a fonte – O autor”

DEDICATÓRIA

Aos meus queridos pais,

Elvira Bressiani de Oliveira e José de Oliveira.

Obrigada pelo carinho, confiança, incentivo e apoio incondicionais, para que meus sonhos se tornassem realidade. Vocês são uma dádiva!

À minha irmã,

Sandra A. Lopes

pela confiança, sempre me desejando êxito.

Às minhas professoras de Matemática do Ensino Fundamental, **Profa. Céris** e **Profa. Néris**. Sou grata por me enxergarem quando eu ainda era “invisível”.

“Os sonhadores agradecem a Deus o espetáculo da vida. Eles não são gigantes nem pessoas especiais, mas pessoas que tombam, choram e se levantam”.

Augusto Cury

“Na essência somos todos iguais, nas diferenças nos respeitamos”.

Stº Agostinho

“Penso noventa e nove vezes, e nada descubro. Deixo de pensar, mergulho no silêncio e a verdade me vem à tona”.

Albert Einstein

“Os que seguem a multidão nunca serão seguidos por ela”.

James Hunter

“Posso fazer qualquer coisa, e se não puder, sei como obter auxílio”.

Ben Carson

“Pensamentos tornam-se ações, ações tornam-se hábitos, hábitos tornam-se caráter, e nosso caráter torna-se o nosso destino”.

James Hunter

“Seja bom com os outros. A distância que você caminha na vida vai depender da sua ternura com os jovens, da sua compaixão com os idosos, de sua compreensão com aqueles que lutam, da sua tolerância com os fortes. Porque algum dia na vida você poderá ser um deles”.

George Washington Carver

AGRADECIMENTOS

A DEUS, doador da vida, da inteligência e de todos os recursos para a realização deste trabalho. Tenho mais do que preciso e mereço. Sou grata pela esperança, felicidade e paz.

À Profa. Roseli Aparecida Leandro pela orientação, ensinamentos, sugestões, críticas e também pela oportunidade de realizar este trabalho.

À Profa. Dra. Anete Pereira de Souza, do Centro de Biologia Molecular e Engenharia Genética da UNICAMP, pela obtenção dos dados genotípicos e moleculares das progênes avaliadas neste trabalho.

Ao querido amigo e Professor Dr. Cláudio Lopes Souza Júnior, do Departamento de Genética da ESALQ/USP, pela co-orientação com seriedade, ética, paciência, firmeza, incentivo e preciosíssimas sugestões que enriqueceram significativamente a qualidade deste trabalho, além do ótimo convívio. Seus ensinamentos foram muito além do “mapeamento de QTLs”. Obrigada pelo exemplo de liderança e de compromisso com a pesquisa.

A todos os alunos do Prof. Cláudio pela amizade sadia e companheirismo, em especial ao Dr. Dyeme A. V. Bento e ao Dr^{ndo} Tassiano M. M. Câmara, pela paciência, atenção e por nunca medirem esforços ao me ensinarem os conceitos genéticos e a trabalhar com o *software* QTLCartographer. Muito obrigada, mesmo!

Aos pesquisadores Dr. Shengchu Wang (Statistical Genetics Department, North Carolina State University) e Dr. Brian S. Yandell (Statistical Department, University of Wisconsin) pelas sugestões sobre a análise Bayesiana no QTLCartographer e no *software* R.

À Profa. Elizabeth Strapasson pelas sugestões a respeito da configuração deste trabalho no editor de texto L^AT_EX.

Aos membros da banca examinadora: Dr. Cláudio Lopes de Souza Jr. e Dr. Luis Aparecido Milan (Departamento de Estatística da UFSCAR-São Carlos) que disponibilizaram tempo lendo e sugerindo alterações para a melhoria deste trabalho.

Aos membros suplentes da banca examinadora: Dra. Sônia Maria De Stefano Piedade (Departamento de Ciências Exatas da ESALQ/USP) e Dr. Julio Sílvio de Sousa Bueno Filho (Departamento de Ciências Exatas da UFLA-Lavras).

À ESALQ/USP, pela infra-estrutura, especialmente aos departamentos de Ciências Exatas e Genética e Melhoramento de Plantas.

Aos professores do Departamento de Ciências Exatas da ESALQ/USP, especialmente ao Prof. Dr. Décio Barbin, pelos conhecimentos científicos e profissionais que muito contribuíram para a minha formação, além do agradável e precioso convívio. Foi muito bom.

Aos funcionários do Departamento de Ciências Exatas da ESALQ/USP, em especial ao Jorge e às queridas secretárias, Luciane e Solange.

Aos colegas de mestrado e doutorado do programa Estatística e Experimentação Agrônômica da ESALQ/USP, em especial o César Taconelli, a Francine Ragonha, o Júlio C. Pereira, Hélio Rubens e Sandra Cardozo, pela convivência sadia durante o curso. Não posso deixar de citar o Elvis França do CENA/USP e o Luis Fernando B. Pinto do Departamento de Zootecnia da ESALQ/USP. Foi um enorme privilégio conhecê-los.

À Esther Rodrigues, pelo agradável convívio em casa. Obrigada pelos momentos de risos, conversas e também pelo silêncio, quando necessário. Você é especial.

Aos sempre amigos Caroline Cremasco, Julianne Matsuoka, Márcio R. Perrigil e Milton Y. Saito, pela sincera e sólida amizade. É muito bom saber que posso contar com vocês.

A todos que, de forma direta ou indireta, contribuíram para a realização deste trabalho.

À CAPES, fundamental suporte financeiro concedido.

SUMÁRIO

	Página
RESUMO	9
ABSTRACT	10
1 INTRODUÇÃO	11
2 DESENVOLVIMENTO	13
2.1 Revisão Bibliográfica	13
2.1.1 Conceitos em Inferência Bayesiana	13
2.1.1.1 O que é Inferência Bayesiana	13
2.1.1.2 Teorema de Bayes	14
2.1.1.3 Função de Verossimilhança	15
2.1.1.4 Distribuições <i>a priori</i>	16
2.1.1.5 Distribuições <i>a priori</i> próprias e impróprias	16
2.1.1.6 Distribuições <i>a priori</i> conjugadas	17
2.1.1.7 Cadeias de Markov	18
2.1.1.8 Métodos de Monte Carlo via Cadeias de Markov (MCMC)	20
2.1.1.9 Intervalo de Credibilidade	22
2.1.1.10 Método de Seleção de Modelos - Fator de Bayes	23
2.1.2 A Importância Econômica do Milho	24
2.1.3 Conceitos Genéticos	25
2.1.4 Interações Alélicas	26
2.1.4.1 Interação Aditiva	27
2.1.4.2 Interação Dominante	27
2.1.4.3 Interação Sobredominante	28
2.1.5 Meiose e <i>Crossing-over</i>	28
2.1.6 Interferência e Coincidência	30
2.1.7 Funções de Mapeamento	31
2.1.8 Marcadores moleculares	33
2.1.9 Desequilíbrio de Ligação	35
2.1.10 Pré-requisitos para o mapeamento de QTLs	35
2.1.10.1 Delineamento genético adequado	36

	8
2.1.10.2 Mapa Genético	36
2.1.10.3 Dados dos caracteres de interesse	38
2.1.11 Mapeamento de QTLs	39
2.1.12 Métodos Estatísticos para o Mapeamento de QTLs	40
2.1.12.1 Análise de Marcas Simples	40
2.1.12.2 Mapeamento por Intervalo (IM)	41
2.1.12.3 Mapeamento por Intervalo Composto (CIM)	47
2.1.12.4 Mapeamento por Intervalo Bayesiano (BIM)	54
2.1.12.5 Especificação das <i>Prioris</i>	58
2.2 Material e Métodos	61
2.2.1 Material genético	61
2.2.2 Mapeamento de QTLs	61
2.2.3 Método 1: Mapeamento por Intervalo Composto	62
2.2.3.1 Estimativas dos parâmetros genéticos	63
2.2.4 Método 2: Mapeamento por Intervalo Bayesiano	65
2.3 Resultados e Discussão	66
2.3.1 Resultados obtidos pelo método 1 (CIM)	66
2.3.2 Resultados obtidos pelo método 2 (BIM)	72
2.3.3 Comparação dos resultados obtidos pelos métodos CIM e BIM	83
3 CONCLUSÕES	85
REFERÊNCIAS	88
ANEXOS	93

RESUMO

Mapeamento de QTLs utilizando as abordagens Clássica e Bayesiana

A produção de grãos e outros caracteres de importância econômica para a cultura do milho, tais como a altura da planta, o comprimento e o diâmetro da espiga, apresentam herança poligênica, o que dificulta a obtenção de informações sobre as bases genéticas envolvidas na variação desses caracteres. Associações entre marcadores e QTLs foram analisadas através dos métodos de mapeamento por intervalo composto (CIM) e mapeamento por intervalo Bayesiano (BIM). A partir de um conjunto de dados de produção de grãos, referentes à avaliação de 256 progênies de milho genotipadas para 139 marcadores moleculares codominantes, verificou-se que as metodologias apresentadas permitiram classificar marcas associadas a QTLs. Através do procedimento CIM, associações entre marcadores e QTLs foram consideradas significativas quando o valor da estatística de razão de verossimilhança (LR) ao longo do cromossomo atingiu o valor máximo dentre os que ultrapassaram o limite crítico $LR = 11,5$ no intervalo considerado. Dez QTLs foram mapeados distribuídos em três cromossomos. Juntos, explicaram 19,86% da variância genética. Os tipos de interação alélica predominantes foram de dominância parcial (quatro QTLs) e dominância completa (três QTLs). O grau médio de dominância calculado foi de 1,12, indicando grau médio de dominância completa. Grande parte dos alelos favoráveis ao caráter foram provenientes da linhagem parental L02-02D, que apresentou mais elevada produção de grãos. Adotando-se a abordagem Bayesiana, foram implementados métodos de amostragem através de cadeias de Markov (MCMC), para obtenção de uma amostra da distribuição *a posteriori* dos parâmetros de interesse, incorporando as crenças e incertezas *a priori*. Resumos sobre as localizações dos QTLs e seus efeitos aditivo e de dominância foram obtidos. Métodos MCMC com saltos reversíveis (RJMC) foram utilizados para a análise Bayesiana e Fator calculado de Bayes para estimar o número de QTLs. Através do método BIM associações entre marcadores e QTLs foram consideradas significativas em quatro cromossomos, com um total de cinco QTLs mapeados. Juntos, esses QTLs explicaram 13,06% da variância genética. A maior parte dos alelos favoráveis ao caráter também foram provenientes da linhagem parental L02-02D.

Palavras-chave: Mapa genético; Mapeamento por intervalo Bayesiano; Mapeamento por intervalo composto; MCMC com saltos reversíveis; Quantitative trait loci; Verossimilhança

ABSTRACT**Mapping QTLs: Classical and Bayesian approaches**

Grain yield and other important economic traits in maize, such as plant height, stalk length, and stalk diameter, exhibit polygenic inheritance, making difficult information achievement about the genetic bases related to the variation of these traits. The number and sites of (QTLs) loci that control grain yield in maize have been estimated. Associations between markers and QTLs were undertaken by composite interval mapping (CIM) and Bayesian interval mapping (BIM). Based on a set of grain yield data, obtained from the evaluation of 256 maize progenies genotyped for 139 codominant molecular markers, the presented methodologies allowed classification of markers associated to QTLs. Through composite interval mapping were significant when value of likelihood ratio (LR) throughout the chromosome surpassed $LR = 11,5$. Significant associations between markers and QTLs were obtained in three chromosomes, ten QTLs has been mapped, which explained 19,86% of genetic variation. Predominant genetic action for mapped QTLs was partial dominance and (four QTLs) complete dominance (tree QTLs). Average dominance amounted to 1,12 and confirmed complete dominance for grain yield. Most alleles that contributed positively in trait came from parental strain L02-02D. The latter had the highest yield rate. Adopting a Bayesian approach to inference, usually implemented via Markov chain Monte Carlo (MCMC). The output of a Bayesian analysis is a posterior distribution on the parameters, fully incorporating prior beliefs and parameter uncertainty. Reversible Jump MCMC (RJMCMC) is used in this work. Bayes Factor is used to estimate the number of QTL. Through Bayesian interval, significant associations between markers and QTLs were obtained in four chromosomes and five QTLs has been mapped, which explained 13,06% of genetic variation. Most alleles that contributed positively in trait came from parental strain L02-02D. The latter had the highest yield rate.

Key words: Genetic mapping; Bayesian interval mapping; Composite interval mapping; Reversible jump MCMC; Quantitative trait loci; Likelihood

1 INTRODUÇÃO

Muitas características animais e vegetais de importância econômica, como altura de plantas, produção de grãos, teor de óleo, dentre outros, apresentam um padrão contínuo de variação, ou seja, não há classes fenotípicas claramente distintas de indivíduos. Essa variação contínua pode ser explicada pela ação conjunta de um grande número de genes - poligenes - que embora apresente um padrão mendeliano normal individualmente, possuem um pequeno efeito sobre o fenótipo, e pelo efeito ambiental (BEARZOTI, 2000).

Este fato constitui-se em uma dificuldade no trabalho de melhoramento, uma vez que a avaliação e seleção de genótipos requer grande precisão experimental e utilização de metodologias estatísticas mais elaboradas. Desse modo, tais caracteres não se prestam para uma análise das causas de variação genética da forma como Mendel mostrou ser possível para os caracteres que apresentam variação discreta ou descontínua.

O desenvolvimento de marcadores genéticos cada vez mais informativos, culminando com o aparecimento dos marcadores moleculares, possibilitou enorme avanço no estudo dos caracteres quantitativos. Os marcadores moleculares aproveitam o polimorfismo da própria molécula de DNA, ocorrendo, portanto, em número suficiente para que seja possível a marcação de qualquer alelo presente em uma dada espécie (RAMALHO; SANTOS; PINTO, 2004). Com isso, foi possível a elaboração de mapas genéticos que apresentassem grande número desses marcadores, etapa fundamental para o início do trabalho de mapeamento de regiões definidas que contêm locos controladores de caracteres quantitativos, denominados QTLs (*Quantitative Trait Loci*).

O procedimento de procura, localização e estimação dos efeitos de QTLs é denominado mapeamento de QTLs. Os delineamentos usuais utilizados nas espécies cultivadas, baseiam-se no uso de populações segregantes, F_2 , ou retrocruzamentos, as quais são obtidas pelo cruzamento de duas linhagens homozigóticas fenotipicamente divergentes para o caráter quantitativo em estudo. Diversas metodologias visando o mapeamento de QTLs foram desenvolvidas a partir do final da década de 1980. Da análise feita inicialmente com marcas simples, seguiu-se o mapeamento por intervalo - IM, proposto por Lander e Botstein (1989), o mapeamento por intervalo composto - CIM, proposto por Zeng (1993; 1994), Jansen e Stam (1993; 1995) e de múltiplos intervalos - MIM Kao, Zeng e Teasdale (1999) sendo observado um contínuo avanço tanto em relação ao número de informações obtidas - presença, posição, efeitos genéticos

e de interações dos QTLs - como em precisão das mesmas, informações de grande importância para o melhoramento.

O presente trabalho visou apresentar e discutir os métodos de mapeamento por intervalo Bayesiano (BIM) baseado na abordagem Bayesiana e mapeamento por intervalo Composto (CIM) baseado na abordagem Clássica, utilizando estimação de máxima verossimilhança e de mínimos quadrados. Considerando-se a abordagem Bayesiana, o número de QTLs foi considerado como uma variável aleatória e, portanto, um dos parâmetros de interesse. Uma amostra da distribuição conjunta *a posteriori* para os parâmetros de interesse foi obtida utilizando métodos Monte Carlo através de Cadeias de Markov com Saltos Reversíveis (RJMCMC). O número de QTLs no modelo foi determinado pelo Fator de Bayes e um conjunto de dados reais referente à produção de grãos de milho foi analisado.

2 DESENVOLVIMENTO

2.1 Revisão Bibliográfica

2.1.1 Conceitos em Inferência Bayesiana

Nesta seção, serão introduzidos alguns conceitos básicos em Inferência Bayesiana, com a finalidade de facilitar a compreensão deste trabalho.

2.1.1.1 O que é Inferência Bayesiana

A noção de informação está presente em todos os estudos desenvolvidos em estatística, pois a estatística trabalha com incertezas. Deste modo, o ideal é ter o máximo possível de informações para decidir com quais se devem trabalhar.

A abordagem estatística Bayesiana considera todas as informações disponíveis, como a opinião de especialistas e experimentos anteriores, úteis para reduzir as incertezas na análise. Em contrapartida, a abordagem Clássica se restringe apenas àquelas informações obtidas com a observação de dados amostrais. Esses pontos de vista formam a base das duas principais abordagens estatísticas: Bayesiana e Clássica.

Em geral, para uma análise estatística são construídos modelos matemáticos, que representam as relações entre as variáveis em estudo. Para a construção de um modelo é necessário fazer com que este contenha o máximo de informações, sem torná-lo muito complexo (parcimônia). Se o modelo for simples demais, não estará informando o suficiente e, caso seja complexo demais, tornará difícil sua compreensão.

A análise Bayesiana para um determinado modelo estatístico, consiste na obtenção da densidade *a posteriori* para os parâmetros de interesse, que são densidades obtidas combinando a informação *a priori* sobre tais parâmetros (densidades *a priori*) e a informação contida na amostra (função de verossimilhança). As vantagens para a utilização desta abordagem estão no fato de não exigir grandes amostras e permitir a inclusão de informações adicionais na análise.

Resumindo, a Inferência Bayesiana requer um modelo amostral (verossimilhança) combinado com uma distribuição *a priori* para os parâmetros, resultando em uma distribuição

conhecida como distribuição *a posteriori*, que resume o resultado sobre os parâmetros do modelo e sobre quantidades não observadas, tais como a predição para novas observações.

2.1.1.2 Teorema de Bayes

O teorema de Bayes, é a chave de toda estatística Bayesiana, pois permite atualizar probabilidades e incorporar um conhecimento prévio sobre parâmetros desconhecidos θ de um modelo probabilístico e atualizar informações continuamente sobre o mesmo, conforme as observações são coletadas.

Suponha que $\mathbf{y}^T = (y_1, \dots, y_n)$ seja um vetor de n observações de uma característica populacional, da qual deseja-se obter inferências sobre uma quantidade de interesse desconhecida não observável, θ , com valores possíveis em um conjunto Θ , definido como espaço paramétrico. A quantidade θ pode ser um escalar, vetor ou matriz. A inferência Bayesiana admite que o parâmetro desconhecido θ da distribuição de \mathbf{Y} pode ser descrito por uma distribuição de probabilidade $\pi(\theta)$, chamada de distribuição *a priori*. Nessa distribuição são incorporadas todas as informações de que se dispõe sobre θ , resumidas probabilisticamente, podendo ser atualizadas à medida que se observa o valor da quantidade aleatória \mathbf{Y} relacionada a θ , isto é, à medida em que os dados y_1, \dots, y_n são observados (PAULINO; TURKMAN; MURTEIRA, 2003).

A distribuição amostral que incorpora a informação sobre o parâmetro fornecida pelos dados observados é a função de verossimilhança, $L(\theta|\mathbf{y})$, que depende de k parâmetros $\theta^T = (\theta_1, \dots, \theta_k)$. A partir das informações *a priori* associadas às informações fornecidas pelas observações, são obtidas as distribuições *a posteriori* de θ , $\pi(\theta|\mathbf{y})$, utilizando o Teorema de Bayes (eq. 1):

$$\pi(\theta|\mathbf{y}) = \frac{f(\mathbf{y}|\theta)}{f(\mathbf{y})} = \frac{L(\theta|\mathbf{y})\pi(\theta)}{f(\mathbf{y})}. \quad (1)$$

em que $f(\mathbf{y}|\theta)$ é a função densidade de probabilidade conjunta, que pode ser escrita como o produto entre a densidade *a priori* e a função de verossimilhança; $f(\mathbf{y})$ é a soma de todos os possíveis valores de θ , quando θ é discreto, ou a integral definida num intervalo de valores possíveis de θ , no caso em que θ seja contínuo.

A função $f(\mathbf{y})$, no denominador, não depende de θ e, portanto, para a determinação da distribuição de interesse $\pi(\theta|\mathbf{y})$, representa apenas uma constante. Por essa razão a forma usual do teorema de Bayes em (1) reduz-se a (eq. 2)

$$\pi(\boldsymbol{\theta}|\mathbf{y}) = cL(\boldsymbol{\theta}|\mathbf{y})\pi(\boldsymbol{\theta}) \quad (2)$$

em que $c^{-1} = f(\mathbf{y})$ é a constante normalizadora, ou ainda, (eq. 3)

$$\pi(\boldsymbol{\theta}|\mathbf{y}) \propto L(\boldsymbol{\theta}|\mathbf{y})\pi(\boldsymbol{\theta}) \quad (3)$$

sendo que o símbolo \propto denota proporcionalidade.

O teorema de Bayes fornece a regra para a atualização da probabilidade sobre $\boldsymbol{\theta}$, partindo de $\pi(\boldsymbol{\theta})$ e chegando a $\pi(\boldsymbol{\theta}|\mathbf{y})$. Por esse motivo, essas distribuições são chamadas de distribuições *a priori* e *a posteriori* (SOUZA, 2003).

2.1.1.3 Função de Verossimilhança

A função de verossimilhança $L(\boldsymbol{\theta}|\mathbf{y})$ desempenha um papel muito importante na fórmula de Bayes. É através dela que os dados podem modificar o conhecimento *a priori* sobre $\boldsymbol{\theta}$. Essa função pode ser vista como a representação do que os dados têm a informar a respeito do parâmetro $\boldsymbol{\theta}$ (LEANDRO, 2001).

Suponha que Y_1, \dots, Y_n é uma amostra aleatória da distribuição $f(\mathbf{y}|\boldsymbol{\theta})$, $\boldsymbol{\theta} \in \Theta$. A função densidade de probabilidade conjunta é

$$f(\mathbf{y}|\boldsymbol{\theta}) = f(y_1|\boldsymbol{\theta})f(y_2|\boldsymbol{\theta})\dots f(y_n|\boldsymbol{\theta}) = \prod_{i=1}^n f(y_i|\boldsymbol{\theta}).$$

Fixando-se o ponto amostral y_1, \dots, y_n a função $L(\boldsymbol{\theta}|\mathbf{y})$, considerada como função de $\boldsymbol{\theta}$, é denominada função de verossimilhança da amostra, dada por

$$L(\boldsymbol{\theta}|\mathbf{y}) = \prod_{i=1}^n f(y_i|\boldsymbol{\theta}).$$

Após a obtenção dos dados, $L(\boldsymbol{\theta}|\mathbf{y})$ pode ser vista como uma função do vetor de parâmetros $\boldsymbol{\theta}$ para um conjunto fixo de dados de valores observados y_1, \dots, y_n , servindo para medir o quanto os dados suportam uma hipótese sobre $\boldsymbol{\theta}$. Para fazer inferência sobre o parâmetro $\boldsymbol{\theta}$ de interesse, baseando-se apenas na função de verossimilhança, é importante somente o que foi realmente observado e não o que “poderia” ter sido mas efetivamente não ocorreu (EHLERS, 2003). A função de verossimilhança $L(\boldsymbol{\theta}|\mathbf{y})$, a partir do teorema de Bayes, é combinada com a distribuição de probabilidade *a priori* para se obter a distribuição de probabilidade *a posteriori*, utilizada para fazer inferências sobre o vetor de parâmetros de interesse.

A verossimilhança não é uma probabilidade. Dessa forma, não faz sentido adicionar verossimilhanças, mas sim efetuar o quociente entre elas. A razão

$$\frac{L(\boldsymbol{\theta}_0|\mathbf{y})}{L(\boldsymbol{\theta}_1|\mathbf{y})}$$

mede o peso da evidência ou plausibilidade de $\boldsymbol{\theta}_0$ contra $\boldsymbol{\theta}_1$ decorrente da observação de y_1, \dots, y_n (PAULINO; TURKMAN; MURTEIRA, 2003).

2.1.1.4 Distribuições *a priori*

A distribuição *a priori* representa informações do parâmetro $\boldsymbol{\theta}$, desconhecido antes da realização do experimento; desempenha um papel importante na inferência Bayesiana. Essa distribuição pode ser utilizada para representar probabilisticamente o conhecimento prévio ou ignorância relativa sobre a quantidade desconhecida $\boldsymbol{\theta}$ (LEANDRO, 2001).

A informação *a priori* que se pretende incorporar na análise é a informação fornecida por um especialista ou pesquisador, tais como dados históricos do problema ou de experimentos análogos (PAULINO; TURKMAN; MURTEIRA, 2003).

A crença *a priori* sobre $\boldsymbol{\theta}$ pode diferir de pesquisador para pesquisador. Há casos em que *a priori* dada é informativa, refletindo algum conhecimento prévio sobre o parâmetro desconhecido. Outras vezes, a distribuição *a priori* é não informativa, permitindo que os dados mostrem as informações sobre $\boldsymbol{\theta}$.

Especificar probabilidades subjetivas que representam as crenças do pesquisador é uma tarefa difícil, pois o ajuste dessa distribuição é feito através do “sentimento” à distribuição *a priori*. Daí percebe-se a importância da interação pesquisador-estatístico, cabendo a este sugerir ao pesquisador medidas que resumam sua subjetividade.

2.1.1.5 Distribuições *a priori* próprias e impróprias

Uma distribuição de probabilidade $f(y|\boldsymbol{\theta})$ é definida como própria quando satisfaz as seguintes propriedades:

$f(y|\boldsymbol{\theta}) > 0$ em todo seu domínio de definição e

$\sum_y f(y|\boldsymbol{\theta}) = 1$, se \mathbf{Y} é variável aleatória discreta,

$\int_y f(y|\boldsymbol{\theta})dy = 1$, se \mathbf{Y} é variável aleatória contínua.

No que se refere a distribuição *a priori*, os conceitos de função de probabilidade (ou função densidade de probabilidade) devem ser estendidos para casos em que, para dados discretos, $\sum_{\Theta} \pi(\theta) = \infty$ e, no caso de dados contínuos, a integral $\int_{\Theta} \pi(\theta)d\theta = \infty$. As funções com tais características são conhecidas como *priori* impróprias.

Ressalta-se, porém, que o caso descrito só é possível para a distribuição *a priori* e, portanto, não deve ser empregado para *a posteriori*, que necessariamente deve satisfazer as propriedades de função densidade de probabilidade (ou função de probabilidade).

2.1.1.6 Distribuições *a priori* conjugadas

Uma família de distribuições *a priori* é conjugada se as distribuições *a posteriori* pertencem à mesma classe de distribuições. Sejam $\mathbf{y}^T = (y_1, \dots, y_n)$ observações de variáveis aleatórias independentes e identicamente distribuídas na família exponencial. Tem-se que a função de distribuição conjunta de $\mathbf{Y}_1, \dots, \mathbf{Y}_n$ pode ser escrita como (LEANDRO, 2001):

$$f(\mathbf{y}|\boldsymbol{\theta}) = \prod_{i=1}^n \exp\{a(\theta)b(y_i) + c(\theta) + d(y_i)\},$$

e sua função de verossimilhança por:

$$L(\boldsymbol{\theta}|y) \propto \exp\{a(\theta) \sum_{i=1}^n b(y_i) + nc(\theta)\}.$$

Supondo-se *a priori* conjugada para $\boldsymbol{\theta}$ dada por:

$$\pi(\boldsymbol{\theta}; k_1, k_2) \propto \exp\{k_1 a(\theta) + k_2 c(\theta)\},$$

obtem-se a seguinte distribuição *a posteriori*

$$\pi(\boldsymbol{\theta}|y) \propto \exp\{a(\theta) [\sum_{i=1}^n b(y_i) + k_1] + c(\theta) [n + k_2]\}.$$

A idéia de distribuição *a priori* conjugada é que as distribuições *a priori* e *a posteriori* pertençam à mesma classe de distribuições e, dessa forma, a atualização do conhecimento que se tem sobre $\boldsymbol{\theta}$ envolve apenas uma mudança nos parâmetros indexadores da família de distribuições *a priori* que difere do parâmetro $\boldsymbol{\theta}$, denominados hiperparâmetros. A partir de algum conhecimento sobre $\boldsymbol{\theta}$, uma família de densidades pode ser definida. Nesse caso, a

distribuição *a priori* é representada por uma forma funcional cujos hiperparâmetros devem ser especificados de acordo com esse conhecimento.

O uso de *prioris* conjugadas é muito importante na estatística Bayesiana, pois o aspecto seqüencial do método Bayesiano pode ser explorado definindo-se apenas uma regra de atualização dos hiperparâmetros já que as distribuições permanecem as mesmas. No entanto, deve-se ter cautela ao trabalhar com *prioris* conjugadas, pois nem sempre elas representam adequadamente a incerteza *a priori*.

2.1.1.7 Cadeias de Markov

A parte dinâmica da teoria das probabilidades na qual se estuda uma coleção de variáveis aleatórias, sob o ponto de vista de sua interdependência e de seu comportamento limite é, geralmente, definida como Teoria dos Processos Estocásticos. Os sistemas que variam de uma forma aleatória através do tempo são representados por modelos matemáticos, conhecidos por **processos estocásticos**.

Um processo estocástico pode ser definido como uma coleção de variáveis aleatórias $\{X_t\}$, $t \in T$, definidas no mesmo espaço de probabilidades, sendo T um subconjunto de $(-\infty, \infty)$ que representa o conjunto de índices ou de parâmetros no espaço de tempo (SOUZA; TACHIBANA, 2001).

Quando T representa um intervalo de comprimento positivo, tem-se um **processo de parâmetro contínuo** e, se T é um subconjunto de inteiros, um **processo de parâmetro discreto**.

O processo estocástico $\{X_n\}$, $n = 1, 2, \dots$ pode assumir um número finito ou infinito enumerável de valores, geralmente denotado pelo conjunto de inteiros não negativos $\{1, 2, \dots\}$. Se $X_n = i$, significa que o processo está no estado x_i , no tempo n . A estrutura mais simples é a independência entre as variáveis, na qual o estado futuro do sistema independe do passado e do presente. No entanto, isso não ocorre na maioria dos sistemas, em que os estados atuais e passados influenciam os estados futuros (PAULINO; TURKMAN; MURTEIRA, 2003). O processo estocástico X_t é uma **cadeia de Markov** se a distribuição de X_t for independente de todos os estados prévios em que se encontrou a cadeia, com exceção do estado imediatamente anterior, isto é (eq. 4),

$$P_{ij} = P[X_t = x_j | X_{t-1} = x_{i_{t-1}}, \dots, X_2 = x_{i_2}, X_1 = x_{i_1}] = P[X_t = x_j | X_{t-1} = x_{i_{t-1}}]. \quad (4)$$

sendo P_{ij} a probabilidade fixa do sistema que em algum momento está no estado x_i partir para outro estado (x_j), denominada **probabilidade de transição**. Quando a probabilidade em (eq. 4) for independente de t , a cadeia será **homogênea** ou seja (eq. 5),

$$P_{ij} = P(x_i \longrightarrow x_j) = P[X_t = x_j | X_{t-1} = x_{i_{t-1}}], \quad (5)$$

satisfazendo $P_{ij} \geq 0$, $i, j > 0$ e $\sum_j^\infty P_{ij} = 1$ para todo i .

A matriz estocástica \mathbf{P} de probabilidades de transição a um passo, construída a partir de P_{ij} é dada por:

$$\mathbf{P} = \begin{bmatrix} P_{00} & P_{01} & P_{02} & \dots \\ P_{10} & P_{11} & P_{12} & \dots \\ \vdots & \vdots & \vdots & \dots \\ P_{i0} & P_{i1} & P_{i2} & \dots \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix}.$$

Ao considerar

$$P_{ij}^{(n)} = P[X_t = x_j | X_{t-n} = x_i],$$

para designar as probabilidades de transição a n passos, ou seja, as probabilidades de, em n passos, partir do espaço x_i para atingir x_j . A matriz $P^{(n)}$ construída a partir de $P_{ij}^{(n)}$, é definida como matriz de probabilidades de transição n passos.

O estado x_j é dito ser **acessível** do estado x_i se $P_{ij}^{(n)} > 0$ para todo $n > 0$. Isto implica que o estado x_j é acessível do estado x_i se e somente se, iniciando em x_i , for possível que processo atinja o estado x_j . Se dois estados x_i e x_j são acessíveis entre si, dizemos que se **comunicam** entre si e podendo-se escrever $x_i \longleftrightarrow x_j$. Se o estado x_i se comunica com o estado x_j e se o estado x_j se comunica com o estado x_k , então o estado x_i se comunica com o estado x_k ($x_i \longleftrightarrow x_k$). Uma cadeia de Markov é **irredutível**, se e somente se, todos os estados se comunicam entre si. A comunicação é uma relação de equivalência e portanto numa cadeia de Markov irredutível todos os estados se encontram na mesma **classe** de estado. Isto é, todos os estados se comunicam entre si.

Diz-se que o estado x_i tem período d se $P_{ii}^{(n)} = 0$ para um n não divisível por d , sendo d o maior inteiro com essa propriedade (isto é, quando for divisível por $P_{ii}^{(n)} > 0$). Por

exemplo, começando-se em x_i o processo volta para o estado x_i , somente nas vezes 6, 12, 18, 24, ..., ou seja, o estado tem período $d = 6$. Se o estado x_i tem período d e se os estados x_i e x_j se comunicam, então o estado x_j também tem período d . Um estado se diz **aperiódico** se tiver período $d = 1$. Uma cadeia de Markov é aperiódica se todos os estados são aperiódicos. Dado que a periodicidade seja uma propriedade de classe, para que uma cadeia de Markov irreduzível seja aperiódica é suficiente que um de seus estados seja aperiódico (PAULINO; TURKMAN; MURTEIRA, 2003).

Um estado x_i é **recorrente positivo** quando, começando por x_i , o tempo esperado até o processo retornar ao estado x_i é finito. Em uma cadeia de Markov de estado finito todos os estados recorrentes são recorrentes positivos. Os estados recorrentes positivos e aperiódicos são denominados **ergódicos** (SOUZA; TACHIBANA, 2001).

2.1.1.8 Métodos de Monte Carlo via Cadeias de Markov (MCMC)

Antes da década de 1990 os métodos Bayesianos, apesar de interessantes, apresentavam poucas aplicações práticas devido à escassez de recursos computacionais, elementos indispensáveis para a abordagem Bayesiana. Todas as áreas da Estatística tiveram um desenvolvimento sem precedentes em anos recentes devido aos métodos computacionalmente intensivos, especialmente aqueles baseados em métodos de Monte Carlo via Cadeias de Markov (MCMC). Algoritmos como o amostrador de Gibbs, Metropolis-Hastings e suas variantes são hoje rotineiramente utilizados para resolver problemas complexos. Existem ainda muitas aplicações interessantes a serem desenvolvidas nas mais variadas áreas, como em modelos de processamento de imagens, em epidemiologia espacial e na análise genética (EHLERS, 2003).

Os métodos MCMC requerem que a cadeia de Markov seja

1. homogênea, isto é, as probabilidades de transição de um estado para outro são invariantes;
2. irreduzível, ou seja, cada estado pode ser atingido a partir de qualquer outro em um número finito de iterações;
3. aperiódica, isto é, não haja estados absorventes.

É importante ressaltar que os valores iniciais influenciam o comportamento da cadeia. No entanto, à medida que o número de iterações aumenta, a cadeia gradualmente “esquece” os valores iniciais e eventualmente converge para uma distribuição $g(\theta)$. Desse modo, é comum descartar as iterações iniciais, como se formassem uma “amostra de aquecimento”.

Como foi dito anteriormente, a estatística Bayesiana é baseada na aplicação do teorema de Bayes, do qual derivam inferências ou decisões relevantes obtidas a partir da distribuição *a posteriori*. Em problemas relativamente simples, tais resultados podem ser adquiridos analiticamente, porém, eles se tornam impossíveis à medida que os problemas se tornam complexos.

Os métodos Monte Carlo via cadeias de Markov (MCMC) são uma importante ferramenta computacional em estatística Bayesiana baseado na simulação de cadeias de Markov irreduzíveis aperiódicas cuja distribuição estacionária $g(\boldsymbol{\theta})$ é a distribuição *a posteriori* de interesse, $\pi(\boldsymbol{\theta}|y)$, da qual não é possível gerar diretamente uma amostra para a obtenção de inferências via Monte Carlo (PAULINO; TURKMAN; MURTEIRA, 2003). Em outras palavras, o método consiste em construir uma cadeia, facilmente simulada, uma “amostra inferencial” muito grande da distribuição *a posteriori*, cujos valores gerados poderão ser usados para resumir características da distribuição *a posteriori*, de interesse (LEANDRO, 2001; O’HAGAN; LUCE; FRYBACK, 2003). Por exemplo, para se estimar a média de uma distribuição *a posteriori*, basta calcular apenas a média desta amostra inferencial. Se a amostra for suficientemente grande, a média da amostra será uma aproximação da média da distribuição verdadeira (EHLERS, 2003).

Esses métodos permitem, através de um processo iterativo, gerar realizações de uma cadeia de Markov de tal modo que, à medida que o número de iterações aumenta, a cadeia se aproxima da sua condição de equilíbrio. Se em determinado instante t a cadeia já se encontrar no estado de equilíbrio, então o vetor de parâmetros $\boldsymbol{\theta}$ gerado nesse instante pode ser considerado como uma realização da distribuição $\pi(\boldsymbol{\theta}|y)$. No entanto, sucessivas realizações de uma mesma cadeia ao longo do tempo, não constituem uma amostra aleatória da distribuição de interesse, pois os vetores $\boldsymbol{\theta}$ que vão sendo gerados são correlacionados. Surge, então, o problema em como utilizar as realizações da cadeia, após o equilíbrio, para fazer inferências sobre $\boldsymbol{\theta}$.

Uma alternativa é gerar uma longa cadeia a partir de um estado inicial, mas usar, para fazer inferências, apenas algumas das observações resultantes. Assim, a partir do estado inicial da cadeia, designado por $\boldsymbol{\theta}^0 = [\theta_1^0, \dots, \theta_k^0]$, gera-se uma longa cadeia de comprimento t , que é o número de iterações, sendo $t = l + k^*m$. O número inicial de iterações necessárias para a cadeia atingir o estágio de equilíbrio, l , é definido por **período de aquecimento**, (“*burn-in*”). Esse período pode ser mais ou menos longo, de acordo com o tipo de problema. O

valor m é o número das iterações que serão utilizadas na aplicação do método Monte Carlo. A escolha de uma dimensão apropriada m para uma amostra depende muito da precisão que se pretende para a estimação. Em diversas aplicações, valores de $m = 1000$ são muito usados. O espaçamento entre as iterações sucessivas, k^* , definido como “*thin*” é utilizado para eliminar a autocorrelação entre as iterações consecutivas (PAULINO; TURKMAN; MURTEIRA, 2003).

Dentre os métodos utilizados para gerar uma cadeia de Markov com distribuição estacionária especificada destacam-se os algoritmos Metropolis-Hastings e amostrador de Gibbs. O algoritmo de Metropolis (METROPOLIS; ROSENBLUTH; TELLER, 1993), há muitos anos é utilizado em Física. Hastings (1970) apresenta uma generalização desse método, que ficou conhecido na literatura por algoritmo Metropolis-Hastings. É usado em situações em que as distribuições condicionais *a posteriori* não apresentam forma conhecida (como a Normal e Gama, por exemplo) e, conseqüentemente, será impossível obter densidades *a posteriori* diretamente a partir dessas distribuições.

Outro método conhecido é o amostrador de Gibbs, *Gibbs Sampler*, desenvolvido por Geman e Geman (1984) como ferramenta para modelos de processamento de imagens. Gelfand e Smith (1990) apresentaram aplicações interessantes em uma grande variedade de problemas estatísticos. O grande interesse por esse amostrador é devido à sua simplicidade conceitual e facilidade de implementação. Esse método consiste em um esquema markoviano de atualização, que permite a obtenção de amostras de uma distribuição conjunta através de amostragens iterativas das distribuições condicionais completas. À medida que o número de iterações aumenta, a cadeia se aproxima de sua condição de equilíbrio. Desse modo, assume-se que a convergência foi atingida em determinada iteração cuja distribuição esteja arbitrariamente próxima da distribuição de equilíbrio.

Uma inovação recente foi a introdução de algoritmos trans-dimensionais, os algoritmos MCMC com saltos reversíveis (RJMCMC) projetados para construir uma cadeia de Markov em que a dimensão do espaço de estados muda ao longo das iterações (GREEN, 1995).

2.1.1.9 Intervalo de Credibilidade

Toda informação obtida a respeito do parâmetro de interesse θ é expressa pela distribuição *a posteriori* e pode ser resumida através de um valor numérico. Porém, é muito importante associar informações sobre quão precisas são as especificações desse valor. Um resumo de $\pi(\theta|y)$ mais informativo do que qualquer estimativa pontual é obtido de uma região

de Θ que contenha uma parte substancial da densidade *a posteriori* (PAULINO; TURKMAN; MURTEIRA, 2003).

Um intervalo (a, b) é chamado intervalo de credibilidade $100(1 - \alpha)\%$ para θ se:

$$\int_a^b \pi(\theta|y)d\theta = 1 - \alpha, \quad (0 \leq \alpha \leq 1).$$

Dada a infinidade de intervalos de credibilidade com o mesmo nível de credibilidade $100(1 - \alpha)$, interessa selecionar aquele que englobe os valores de θ mais plausíveis *a posteriori*, isto é, selecionar a região que não exclua nenhum valor mais provável do que aqueles nela incluídos (BOX; TIAO, 1992). Essa região é denominada região HPD (“*Highest Posterior Density*”).

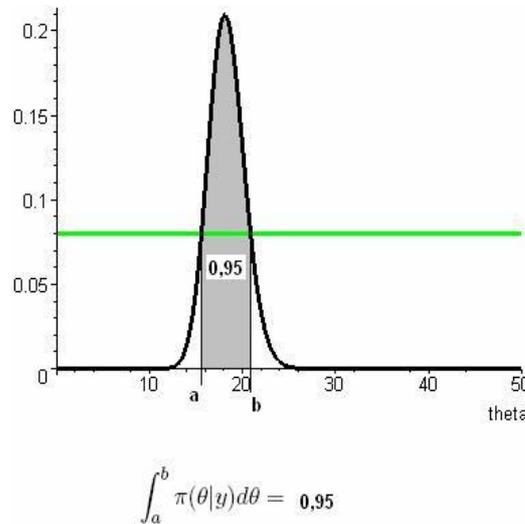


Figura 1 - Intervalo de Credibilidade de 95% para θ

Quanto menor for a amplitude do intervalo mais concentrada está a distribuição do parâmetro, ou seja, a amplitude do intervalo informa sobre a dispersão de θ .

2.1.1.10 Método de Seleção de Modelos - Fator de Bayes

O Fator de Bayes é um método utilizado para selecionar e comparar modelos.

Suponha que duas hipóteses H_j e $H_{(j-1)}$, traduzidas por dois modelos M_j e $M_{(j-1)}$ devam ser verificadas. O Fator de Bayes (BF) é definido por (eq. 6):

$$BF(M_j, M_{(j-1)}) = \frac{P[M_j|\mathbf{y}]/P[M_j]}{P[M_{(j-1)}|\mathbf{y}]/P[M_{(j-1)}]} \quad (6)$$

sendo $P[M_j|\mathbf{y}]$ e $P[M_{(j-1)}|\mathbf{y}]$ as probabilidades *a posteriori* condicionadas às observações para os modelos M_j e $M_{[j-1]}$ respectivamente e, $P[M_j]$ e $P[M_{(j-1)}]$, as probabilidades *a priori* para

os modelos M_j e $M_{(j-1)}$.

Geralmente, não há, por parte do pesquisador, um favoritismo *a priori* por um ou outro modelo antes da realização do experimento. Por essa razão, o Fator de Bayes auxilia na conclusão sobre qual é o modelo mais adequado. Em 1935, Jeffreys (apud RAFTERY, 1995) foi o primeiro a utilizar o fator de Bayes na comparação de teorias científicas, criando inclusive uma regra para a tomada de decisões em função do seu valor. A Tabela 1 apresenta uma calibração para o Fator de Bayes baseado no valor de $B(M_j, M_{(j-1)})$.

Tabela 1 - Decisões sobre a evidência de M_j em relação a $M_{(j-1)}$

Valores de $B(M_j, M_{(j-1)})$	Conclusão
$1 \leq B(M_j, M_{(j-1)}) \leq 3$	Evidência a favor de M_j
$3 < B(M_j, M_{(j-1)}) \leq 10$	Evidência positiva a favor de M_j
$10 < B(M_j, M_{(j-1)}) \leq 100$	Forte evidência a favor de M_j
$B(M_j, M_{(j-1)}) > 100$	Evidência decisiva a favor de M_j

Fonte: JEFFREYS, 1935, apud RAFTERY, 1995.

Em situações envolvendo modelos com muitos parâmetros, o cálculo algébrico do fator de Bayes torna-se inviável. Entretanto, os métodos computacionais atuais contornam esse problema, possibilitando seu cálculo aproximado.

2.1.2 A Importância Econômica do Milho

A utilização do milho na alimentação humana remonta há séculos, constituindo um alimento tradicional da dieta de vários povos, principalmente aqueles que se originaram das civilizações Asteca, Maia e Inca. O seu valor nutritivo, diretamente relacionado aos compostos químicos presentes nos grãos, é um dos principais motivos da sua utilização tanto na alimentação humana quanto animal. A espécie *Zea mays L.* ocupa um destaque no mundo, pelo seu valor econômico, como também à sua capacidade adaptativa. Além do uso direto na propriedade agrícola, o milho apresenta grande versatilidade de aplicação, podendo ser utilizado na produção de ração para aves e suínos; dos grãos se extraem óleo, fubá, farinha, canjica, flocos, amido; das palhas se fazem forragens para animais e peças de artesanato; o “cabelo de milho” é amplamente utilizado na medicina popular como diurético; e o sabugo, também usado para a confecção de placas isolantes.

O alto valor nutritivo do milho é devido à presença nas sementes de grandes

quantidades de carboidratos, proteínas, óleo e vitamina (GARCIA, 2003). O principal consumo do milho ocorre na alimentação animal, por ser um elemento essencial nas rações balanceadas, absorvendo quase a metade do volume total de milho produzido. Portanto, a quantidade de milho está intrinsecamente ligada às cadeias pecuaristas, mais precisamente, a avicultura e a suinocultura (LIMA, 2006).

A maioria dos processos metabólicos afeta a produção da planta de milho. Por essa razão, a herança da produção de grãos envolve diversos aspectos genéticos, tais como os que controlam peso de grão, número de fileiras de grãos na espiga, número de grãos por fileira, comprimento e diâmetro da espiga, fotossíntese, desenvolvimento da raiz, tempo de germinação, entre outros. Por essa razão, esse caráter é um dos mais complexos dentre os avaliados no mapeamento que, além disso, é o mais importante para os melhoristas (BENTO, 2006).

Devido ao fato de a variabilidade da produtividade de grãos de milho ser controlada por diversos locos, o caráter é considerado poligênico. Com o advento dos marcadores moleculares tornou-se possível mapear regiões cromossômicas que afetam caracteres quantitativos, como a produtividade de grãos. Dessa forma, essas regiões denominadas QTLs, podem ser monitoradas e pesquisadas, obtendo-se como consequência maior conhecimento do controle genético do caráter e de sua manipulação no melhoramento (SOUZA JR., 1992, apud MANGOLIM, 2002).

2.1.3 Conceitos Genéticos

Nesta seção, serão introduzidos alguns conceitos básicos em Genética, com a finalidade de facilitar a compreensão deste trabalho.

Conceitos Básicos

O primeiro princípio da genética se refere ao fato de que o fenótipo é resultante do genótipo (G), do ambiente (A) e da interação destes (GA). (GRIFFITHS et al., 2002). Assim, tem-se (eq. 7):

$$F = G + A + GA \quad (7)$$

em que F é o fenótipo, ou seja, formas alternativas de expressão de uma característica; G é o genótipo ou constituição genética de um indivíduo; A é o ambiente ou o conjunto das condições

externas ao organismo e que afetam o seu crescimento e desenvolvimento e GA é a interação entre o genótipo e o ambiente.

Os valores F , G , A e GA podem ser expressos em quaisquer unidades que representem uma propriedade biológica que possa ser medida de maneira contínua, como peso, altura, teor de óleo, produção de grãos etc. Tais caracteres, denominados **quantitativos**, não apresentam classes fenotípicas claramente distintas (COELHO; SILVA, 2005).

Os **poligenes** são conjuntos de genes que participam da manifestação fenotípica de determinado caráter, situando-se em locais nos cromossomos denominados **locos**. Quando estão em um mesmo loco em cromossomos homólogos, tais genes assumem formas alternativas responsáveis pelas diferentes manifestações fenotípicas do caráter e são denominados **alelos**. Uma população pode conter um gene que apresente, por exemplo, dois alelos: A e a . Quando um indivíduo tem a combinação de duas cópias A , então seu genótipo é AA e ele é **homozigoto** para o alelo A . Um indivíduo com genótipo é Aa é **heterozigoto** e um indivíduo com genótipo aa é **homozigoto** para o alelo a . Se cada um dos três genótipos (AA , Aa , aa) determinarem três fenótipos distintos, então os alelos são ditos **codominantes**. Entretanto, se os indivíduos com genótipos AA e Aa expressarem o mesmo fenótipo, então A pode ser definido como **alelo dominante** e a como **alelo recessivo**.

2.1.4 Interações Alélicas

A ação combinada dos alelos de um mesmo gene em um heterozigoto que resulta na expressão do seu fenótipo é denominada interação alélica, a qual pode ser aditiva, dominante ou sobredominante (FALCONER; MACKAY, 1996).

A Figura 2 mostra os valores genotípicos para determinado loco A , em que μ representa o ponto médio entre os dois genótipos homozigóticos, a mede o afastamento de cada genótipo homozigoto em relação à média, e d mede o afastamento do heterozigoto em relação à média μ , isto é, a contribuição dos homozigotos é fornecida por $\pm a$ e do heterozigoto por d , em relação à μ (RAMALHO; SANTOS; PINTO, 2004).

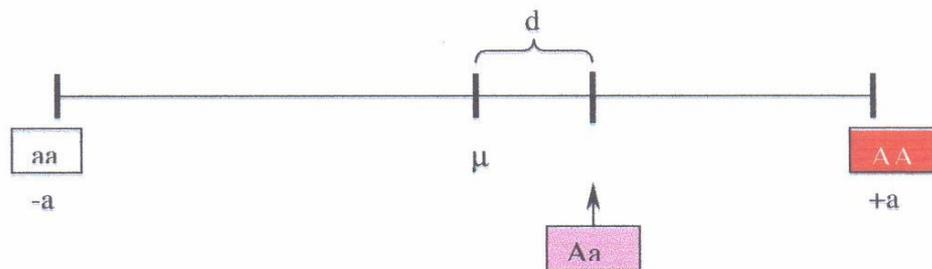


Figura 2 - Valores genotípicos para o loco A

Através dos desvios a e d , os diferentes tipos de interação alélica podem ser avaliados. Se $d = 0$, não há dominância e a interação alélica é denominada aditiva; $d = a$ indica interação alélica de dominância completa; se $0 < d < a$ a interação é de dominância parcial e quando $d > a$ ocorre sobredominância. A relação $|d|/|a|$ mede o grau de dominância de um gene, cujo valor dá uma idéia do tipo de interação alélica geral para o caráter estudado. Quando este valor é igual a zero, a interação alélica é aditiva; se for 1,0, há dominância completa; para valores compreendidos entre zero e um, ocorre dominância parcial e acima de um, há sobredominância.

A Tabela 2 apresenta, resumidamente, a escala dos valores genotípicos.

Tabela 2 - Tipos de Interações Alélicas

$ d / a = 0$	$ d / a = 1$	$0 < d / a < 1$	$ d / a > 1$
Aditiva	Dominância Completa	Dominância Parcial	Sobredominância

2.1.4.1 Interação Aditiva

Na ocorrência de interação alélica aditiva, cada alelo contribui com um efeito fenotípico, que é somado aos efeitos dos demais alelos. A média da primeira geração, (F_1), é igual à média dos genitores, P_1 e P_2 respectivamente, ou seja (eq. 8):

$$\overline{F_1} = \frac{\overline{P_1} + \overline{P_2}}{2}. \quad (8)$$

2.1.4.2 Interação Dominante

Quando a interação alélica é dominante, a contribuição será de cada loco e não de cada alelo. A média da geração F_1 , pode ser igual ao valor de um dos pais, porém será sempre diferente da média dos genitores, isto é, média da F_1 é igual a média da contribuição do genitor

que apresentar maior valor. Quando ocorre dominância completa, a média da geração F_1 pode superar a média da contribuição do genitor superior. Desse modo, tem-se (eq. 9):

$$\overline{F_1} \geq \overline{P}_{superior}. \quad (9)$$

2.1.4.3 Interação Sobredominante

Quando a interação alélica é por sobredominância, no caso de genitores contrastantes a média da geração F_1 será superior à contribuição do genitor superior (10).

$$\overline{F_1} > \overline{P}_{superior}. \quad (10)$$

2.1.5 Meiose e *Crossing-over*

A meiose é o processo de divisão celular que origina os gametas. Estes gametas transmitem os alelos dos genitores para os descendentes na reprodução sexual. Nessa divisão, ocorrem vários fenômenos, que se constituem em fundamentos de diversos tópicos da genética.

A meiose permite a recombinação dos genes - aparecimento de combinações novas não existentes entre os genitores o que, evidentemente, contribui para ampliar a variabilidade genética, principal matéria prima para os melhoristas de plantas e animais (RAMALHO; SANTOS; PINTO, 2004).

Durante a meiose, partes de segmentos homólogos de cromátides não irmãs se ajustam em pares homólogos, e em seguida pode ocorrer troca de material genético, recombinando, assim, os cromossomos (Figura 3). Essa troca é denominada de *crossing-over* ou permuta genética e depende da distância entre os locos. O resultado dessa combinação é o aparecimento de gametas não parentais (recombinantes), a uma determinada frequência ou taxa de recombinação r . Quando os genes estão muito próximos no cromossomo, a separação deles não se realiza e, neste caso, diz-se que ocorre ligação completa. Quando há permutação genética, ocorre ligação parcial. Os genes que segregam independentemente na meiose são ditos não ligados.

Espera-se que o número máximo de recombinantes existentes seja a metade do número de gametas possíveis, isto é, considerando que durante a meiose ocorreu 100% de *crossing-over*, 50% dos gametas formados serão parentais e os outros 50% serão recombinantes,

de modo que a frequência de recombinação é sempre menor do que 50%, ou seja, $r \leq 0,5$. Isso deve-se ao fato de que, sendo um cromossomo formado por duas cromátides, conhecidas como cromátides irmãs, durante a meiose apenas uma delas troca material genético com a cromátide do seu homólogo. Logo, as outras duas cromátides que não trocaram material genético resultam nos gametas parentais. Vale ressaltar que, para a recombinação se manifestar, é preciso que ocorra um número *ímpar* de permutas entre os locos, pois com a ocorrência de um número par de permuta genética a configuração original é restaurada (BEARZOTI, 2000).

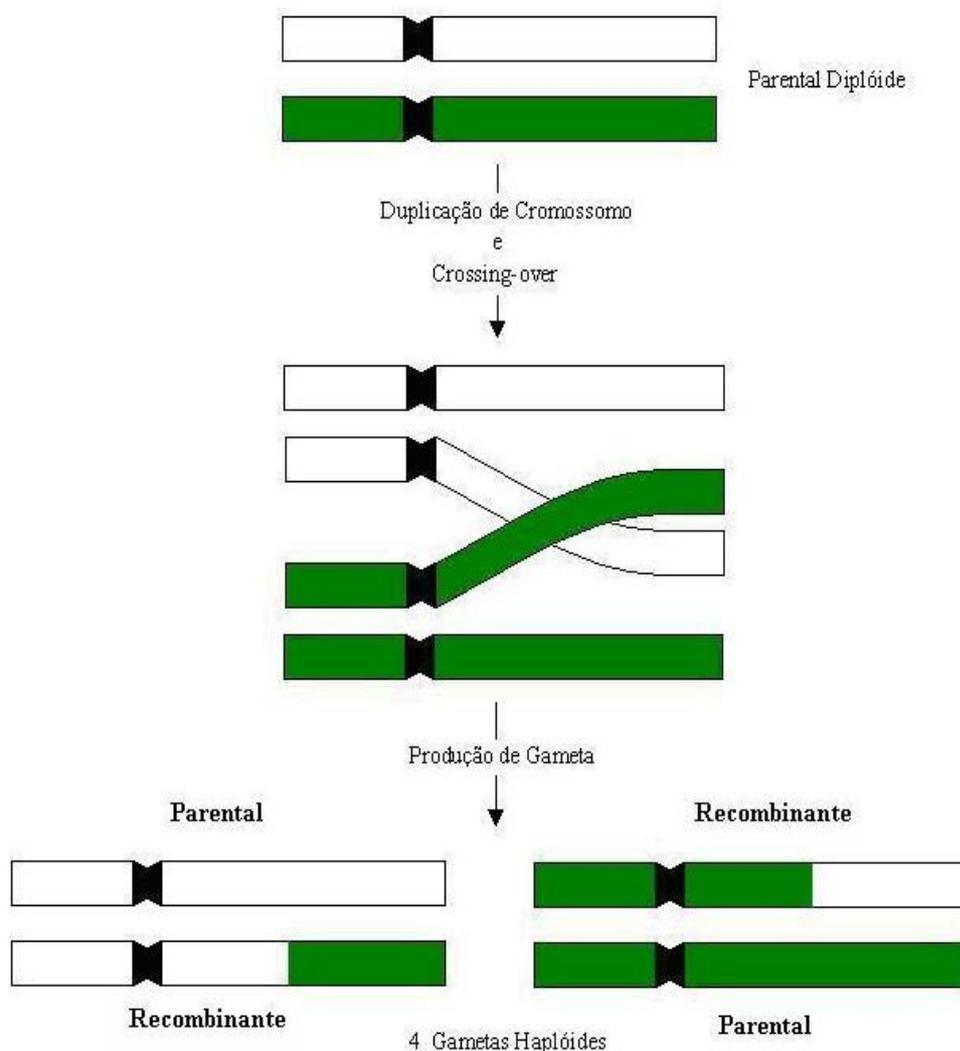


Figura 3 - Recombinação entre um par de cromossomos homólogos

Note que a permuta envolve somente duas das quatro cromátides do par de cromossomos homólogos (Figura 3). Essas duas cromátides trocam segmentos correspondentes por mecanismos de quebra e troca. Observe também que dos quatro produtos desse evento meiótico somente dois contêm recombinações dos alelos dos dois genes. Os outros dois cromossomos filhos carregam combinações parentais.

2.1.6 Interferência e Coincidência

Conceitos de interferência e coincidência são de extrema importância para a construção de mapa genético e para o mapeamento de QTLs.

Seja, por exemplo, um cromossomo com três locos M_1 , M_2 e M_3 , em que a região I representa o segmento M_1M_2 , e a região II o segmento M_2M_3 (Figura 4).

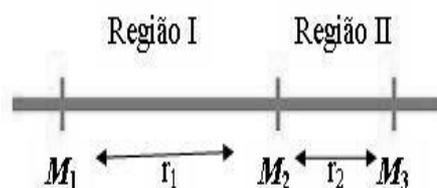


Figura 4 - Representação de três locos em um cromossomo

Através da lei do produto de probabilidades, quando a ocorrência de permuta genética (ou *crossing-over*) em uma determinada região independe da ocorrência de permuta genética em outra, a frequência de permutas duplas entre os locos M_1 e M_3 deveria ser igual ao produto entre as frequências de permuta genética das regiões I (r_1) e II (r_2). Isso é verdade somente se as distâncias que envolvem as duas regiões forem grandes, como por exemplo, situações envolvendo genes situados em extremidades opostas dos cromossomos.

Quando uma permuta ocorrida pode reduzir a probabilidade da ocorrência de permuta genética em região adjacente, tal fenômeno é chamado de interferência. Deste modo, quando a distância entre os genes em um par de cromossomos homólogos influencia na ocorrência de permuta genética, sua probabilidade não será a mesma em todas as partes do cromossomo. Espera-se, portanto, que a interferência também seja desigual em diferentes partes do cromossomo.

Os graus de interferência podem ser medidos através da coincidência, ou seja, pode-se por exemplo, verificar se a ocorrência de permuta genética na região I interfere na ocorrência de permuta na região II, isto é, se a recombinação entre M_1 e M_3 pode ser afetada por M_2 .

A interferência é complementar à coincidência. À medida que a interferência diminui, a coincidência aumenta. Quando a coincidência assume valor 1, conclui-se que a

ocorrência de permuta na região I não interfere na ocorrência de permuta genética na região II. Em contrapartida, no caso em que a coincidência é nula, conclui-se que, com a ocorrência de permuta na região I, será afetada a ocorrência de permuta na região II (RAGONHA, 2005).

2.1.7 Funções de Mapeamento

Um problema encontrado ao se utilizar a frequência de recombinação, r , como medida de distância entre genes para a construção de mapas de ligação, consiste no fato dessa não ser uma medida aditiva (SILVA, 2001). Essa falta de aditividade entre pares de marcas expressas pela frequência de recombinação levou ao desenvolvimento das funções de mapeamento, utilizadas para converter as frequências de recombinação, expressas em unidades de mapa (cM =centiMorgan), em medidas de distâncias com propriedades interessantes para o ordenamento dos locos. As mais conhecidas são as de Haldane e de Kosambi. A função de Haldane admite a independência das permutas nos intervalos adjacentes, tornando os dados de distâncias entre locos (genes ou marcadores) aditivos. A função de Kosambi, por sua vez, considera a interferência nesses intervalos (SCHUSTER; CRUZ, 2004).

Para a solução do problema da falta de aditividade foi proposta outra escala (λ), definida como o número médio de pontos de permuta no segmento em questão. A unidade dessa escala foi denominada Morgan (M), de forma que um Morgan ($1M$) corresponde a um segmento cromossômico no qual ocorre, em média, uma permuta genética por gameta em cada geração, sendo muito comum utilizar a unidade cM (centiMorgan). Na prática, o que interessa, ao se consultar um mapa de ligação, é a frequência de recombinação (r) entre dois locos. Desse modo, é necessário haver uma função de mapeamento, ou seja, uma relação que permita obter r a partir de λ (BEARZOTI, 2000).

Para obter a função Haldane de mapeamento, são admitidas duas pressuposições:

1. As ocorrências de permuta genética são independentes (não há interferência);
2. As ocorrências de permuta genética são aleatórias, ao longo do cromossomo.

Assim, o número de permutas genéticas pode ser modelado por uma distribuição de Poisson.

Para a recombinação entre dois locos separados por uma distância λ se manifestar, é preciso que ocorra um número *ímpar* de permutas entre os locos, pois com a ocorrência de um número par de permuta genética a configuração original é restaurada. Portanto, a frequência r

de recombinação corresponde à probabilidade de ocorrência de um número ímpar de permutas genéticas no intervalo λ . Como o número de *crossing-over*, em um segmento de tamanho λ , pode ser modelado segundo a distribuição de Poisson, com média λ , tem-se:

$$\begin{aligned}
 r &= P(\text{Número ímpar de permutas genéticas}) \\
 &= \sum_y \frac{\lambda^y e^{-\lambda}}{y!} \\
 &= e^{-\lambda} \left[\frac{\lambda}{1!} + \frac{\lambda}{3!} + \frac{\lambda}{5!} + \dots \right] \\
 &= e^{-\lambda} \frac{(e^\lambda - e^{-\lambda})}{2} \\
 &= \frac{1}{2}(1 - e^{-2\lambda})
 \end{aligned} \tag{11}$$

cuja função inversa é:

$$\lambda = -\frac{1}{2} \ln(1 - 2r), \tag{12}$$

conhecida como função de Haldane.

A função de Kosambi, como visto, admite interferência, ou seja, que a possibilidade de ocorrências de permutas genéticas não sejam eventos independentes.

Para estabelecer um modelo que leve em conta o fenômeno da interferência, considere o segmento cromossômico com três genes M_1, M_2 e M_3 e dois segmentos (Figura 4). Desse modo, a probabilidade de permuta entre M_1 e M_3 ignorando M_2 , aqui representada por r_{12} , admitindo ausência de interferência será igual à probabilidade $P[R_I \cap R_{II}^c]$ da ocorrência de recombinação na região I (segmento M_1M_2) mas de não ocorrência na região II (segmento M_2M_3), **ou** a probabilidade $P[R_I^c \cap R_{II}]$ da ocorrência de recombinação na região II (segmento M_2M_3) mas de não ocorrência na região I (segmento M_1M_2). Admitindo que as ocorrências de permutações genéticas são eventos independentes entre si, tem-se

$$P \left[\left(R_I \cap R_{II}^c \right) \cup \left(R_I^c \cap R_{II} \right) \right] = P(R_I) + P(R_{II}) - 2P \left(R_I \cap R_{II} \right) \tag{13}$$

pois, da teoria das Probabilidades,

$$P(R_I \cap R_{II}^c) = P(R_I) - P(R_I \cap R_{II}^c) \text{ e}$$

$$P(R_I^c \cap R_{II}) = P(R_{II}) - P(R_I \cap R_{II}^c),$$

lembrando que $(R_I \cap R_{II}^c)$ e $(R_I^c \cap R_{II})$ são mutuamente exclusivos, R_I e R_{II} são as regiões I

e II representadas na Figura 4.

Portanto, substituindo em (13) os valores fornecidos pela Figura 4 tem-se (eq. 14):

$$r_{12} = r_1 + r_2 - 2r_1r_2. \quad (14)$$

Desvios em relação aos valores obtidos pela eq. 14 têm sido observados, atribuídos ao fenômeno da interferência (LIU, 1998). Desse modo, um modelo geral estabelece que (eq. 15):

$$r_{12} = r_1 + r_2 - 2(1 - \delta)r_1r_2. \quad (15)$$

sendo δ , por definição, um parâmetro entre 0 e 1 que quantifica a interferência. Se $\delta = 0$, remete-se ao modelo de Haldane que admite ausência de interferência, indicando que a ocorrência de permutas genéticas são independentes. Quando $\delta = 1$ se refere à interferência completa, ou seja, a probabilidade de ocorrer permuta em uma região impede a probabilidade da ocorrência de permuta em outra.

Existem várias funções de mapeamento capazes de contemplar o fenômeno da interferência, sendo a função de mapeamento de Kosambi a mais utilizada pelos geneticistas (LYNCH; WALSH, 1998):

$$r = \frac{1}{2} \left(\frac{e^{4\lambda} - 1}{e^{4\lambda} + 1} \right) = \frac{1}{2} \tanh 2\lambda, \quad (16)$$

em que \tanh é a função tangente hiperbólica, cuja função inversa na qual para um dado r pode-se calcular o valor de λ correspondente é dada por (eq. 17):

$$r = \frac{1}{4} \ln \left(\frac{1 + 2r}{1 - 2r} \right), \quad (17)$$

conhecida como função de Kosambi.

2.1.8 Marcadores moleculares

Uma vez que o efeito individual de cada QTL não pode ser observado diretamente nos indivíduos de uma população, uma possibilidade seria identificá-los indiretamente, através

de marcas alocadas nos cromossomos sem efeito sobre os fenótipos dos indivíduos. Essas marcas são denominadas marcadores moleculares. Os marcadores são utilizados para construir os mapas genéticos, necessários para o mapeamento de QTLs (BENTO, 2006; GARCIA, 2003).

Como dito anteriormente, os caracteres quantitativos, não apresentam classes fenotípicas distintas, devido à variação contínua, explicada pela ação conjunta dos poligenes e dos efeitos ambientais.

Com o desenvolvimento das técnicas modernas de biologia molecular, tornou-se possível detectar a variação genética diretamente no DNA. Esse fato foi importante, pois um grande número de marcadores moleculares tornou-se disponível para as espécies de interesse econômico.

O desenvolvimento de marcadores isoenzimáticos proporcionou um enorme avanço no estudo de mapeamento de QTLs. Os marcadores isoenzimáticos apresentam maior polimorfismo em relação aos marcadores fenotípicos, são de fácil avaliação, possuem normalmente controle genético simples e são de baixo custo. Entretanto, tais marcadores não são encontrados em grande número no genoma.

Os marcadores moleculares começaram a ser utilizados na década de 1980, e mais de uma dezena de procedimentos já foram desenvolvidos. Essa classe de marcadores utiliza o polimorfismo da molécula do DNA, o que assegura um número de marcas suficiente para identificar quaisquer alelos da espécie em estudo (RAMALHO; SANTOS; PINTO, 2004).

A classe mais polimórfica de marcadores moleculares utilizada atualmente é a dos SSR (*Simple Sequence Repeats*) ou microssatélites. Os microssatélites são abundantes e distribuídos ao acaso no DNA de animais e plantas, o que permite uma cobertura completa de quaisquer desses genomas. Esses marcadores são altamente polimórficos, estáveis, codominantes e herdáveis de maneira mendeliana (MORGANTE; OLIVIERI, 1993, apud BENTO, 2006). Estas características fazem dos microssatélites os marcadores mais apropriados para o mapeamento genético.

A vantagem do emprego dos marcadores microssatélites, tendo em vista os vários tipos existentes, está no fato de possuir algumas características específicas, fundamentais para seu uso no mapeamento de QTLs. Para que um marcador seja considerado ideal, ele deve ser altamente polimórfico, cobrir todo o genoma, ser seletivamente neutro, não afetando o caráter e nem as características reprodutivas do indivíduo, além de ser codominante, de forma que todos os possíveis locos marcadores sejam identificados e apresentar segregação mendeliana.

A introdução da teoria para interpretar a ligação entre marcador e QTLs, juntamente com o desenvolvimento de marcadores moleculares, que permitem detectar variação genética diretamente no DNA, contribuíram para a elaboração de métodos de análise baseados em marcadores.

2.1.9 Desequilíbrio de Ligação

O desequilíbrio de ligação é definido como qualquer desvio das frequências alélicas esperadas em relação às frequências esperadas de independência, indicando a existência de uma associação entre dois locos. Assim, para que os marcadores moleculares possam contribuir nos estudos de herança e no melhoramento, é necessário que a população esteja em desequilíbrio de ligação, pois, caso contrário, a probabilidade de ocorrência de determinada classe de marcador seria independente da segregação dos alelos do QTL em questão. Em vista disso, é que se utilizam os cruzamentos controlados, chamados de delineamentos genéticos, que visam maximizar o desequilíbrio de ligação.

A detecção de ligação genética entre marcadores e QTLs depende da existência de desequilíbrio de ligação entre os alelos do loco marcador e do QTL. Esse desequilíbrio gera efeitos quantitativos associados ao marcador, que podem ser detectados e estimados por meio de análises estatísticas adequadas.

2.1.10 Pré-requisitos para o mapeamento de QTLs

O mapeamento de QTLs não se trata apenas de determinar a posição destes no genoma, mas também a quantificação e a caracterização de seus efeitos, como interações alélicas - por exemplo, o grau de dominância - de cada um, e mesmo interações epistáticas e pleiotrópicas. (BEARZOTI, 2000). Uma vez que os genótipos desses locos controladores de caracteres quantitativos são desconhecidos, uma alternativa é estimar os efeitos de QTLs fixados em posições intermediárias de marcadores moleculares genotipados que não têm efeito sobre o caráter em estudo, utilizando a informação do mapa desses marcadores.

Os pré-requisitos para o mapeamento de QTLs são:

1. Delineamento genético adequado: Populações do tipo F_2 , de retrocruzamentos, linhagens recombinantes, dentre outras.
2. Genotipagem das plantas

3. Mapa genético

4. Dados dos caracteres de interesse

2.1.10.1 Delineamento genético adequado

Um fator muito importante para a precisão e qualidade do mapeamento de QTLs é o tipo de delineamento genético considerado. As populações derivadas de cruzamentos controlados são interessantes devido à flexibilidade na escolha dos pais na determinação do esquema de cruzamento, pois além da variabilidade encontrada nos genes controladores de um caráter e do ambiente, as variações genéticas do caráter entre os pais devem ser consideradas. São usuais os delineamentos de Populações F_2 , de Retrocruzamentos, Linhagens recombinantes etc. Um dos delineamentos mais utilizados na construção de mapas genéticos vegetais é representado pela população de indivíduos da geração F_2 , resultante do cruzamento entre duas linhagens homocigotas contrastantes para os caracteres de interesse, que será descrito a seguir.

Efetuando o cruzamento entre as linhagens contrastantes, obtém-se a geração F_1 . A seguir, são selecionados em F_1 's os marcadores polimórficos com base na presença de alelos parentais, e em seguida a sua autofecundação, que dará origem à geração F_2 . Posteriormente, é feita a genotipagem das plantas da população F_2 e dentre elas são escolhidas quais serão autofecundadas, gerando progênies $F_{2:3}$ que serão avaliadas em experimentos com repetições, usualmente em diversos ambientes.

A principal vantagem em se utilizar esse tipo de delineamento está na maximização da variabilidade genotípica, pois são obtidos três genótipos para cada loco marcador. Isso permite estimar o grau médio de dominância associado ao QTL detectado, cuja estimativa está diretamente associada a uma escolha eficiente nos métodos de melhoramento a serem adotados na cultura. Além disso, o delineamento F_2 garante o desequilíbrio de ligação, fator indispensável para o mapeamento de QTLs.

2.1.10.2 Mapa Genético

O mapeamento genético de organismos superiores é possível porque o genoma é organizado e transmitido como unidades lineares denominadas cromossomos. Os mapas genéticos apresentam a ordem dos locos ao longo do cromossomo e a distância relativa entre eles. Estes mapas são essenciais para a localização de QTLs.

Para a construção do mapa genético devem ser estabelecidas tanto a posição dos

marcadores como a distância entre eles (LYNCH; WALSH, 1998). Os mapas são construídos com base em análises de segregação dos marcadores em populações que apresentam desequilíbrio de ligação.

Não existe uma relação universal entre distância de mapa e a distância física entre locos. Essa relação dentro de um mesmo cromossomo pode sofrer uma diferença marcante, principalmente próximo aos telômeros e centrômeros. A taxa de recombinação está sob controle genético e alguns genes influenciam essa taxa ao longo de todo o genoma e outros em regiões específicas do cromossomo.

O número necessário de marcadores para construir um mapa genético depende do tamanho do genoma, do número de cromossomos e da frequência de recombinação. Um mapa pode ser considerado satisfatório quando o número de grupos de ligação obtidos pela análise dos marcadores for igual ao número de cromossomos gaméticos do organismo e quando todos os marcadores genéticos mapeados estiverem ligados, indicando que todas as regiões do genoma estão representadas. A Figura 5 é exemplo de um mapa genético dos 10 cromossomos de milho construído com 139 marcas.

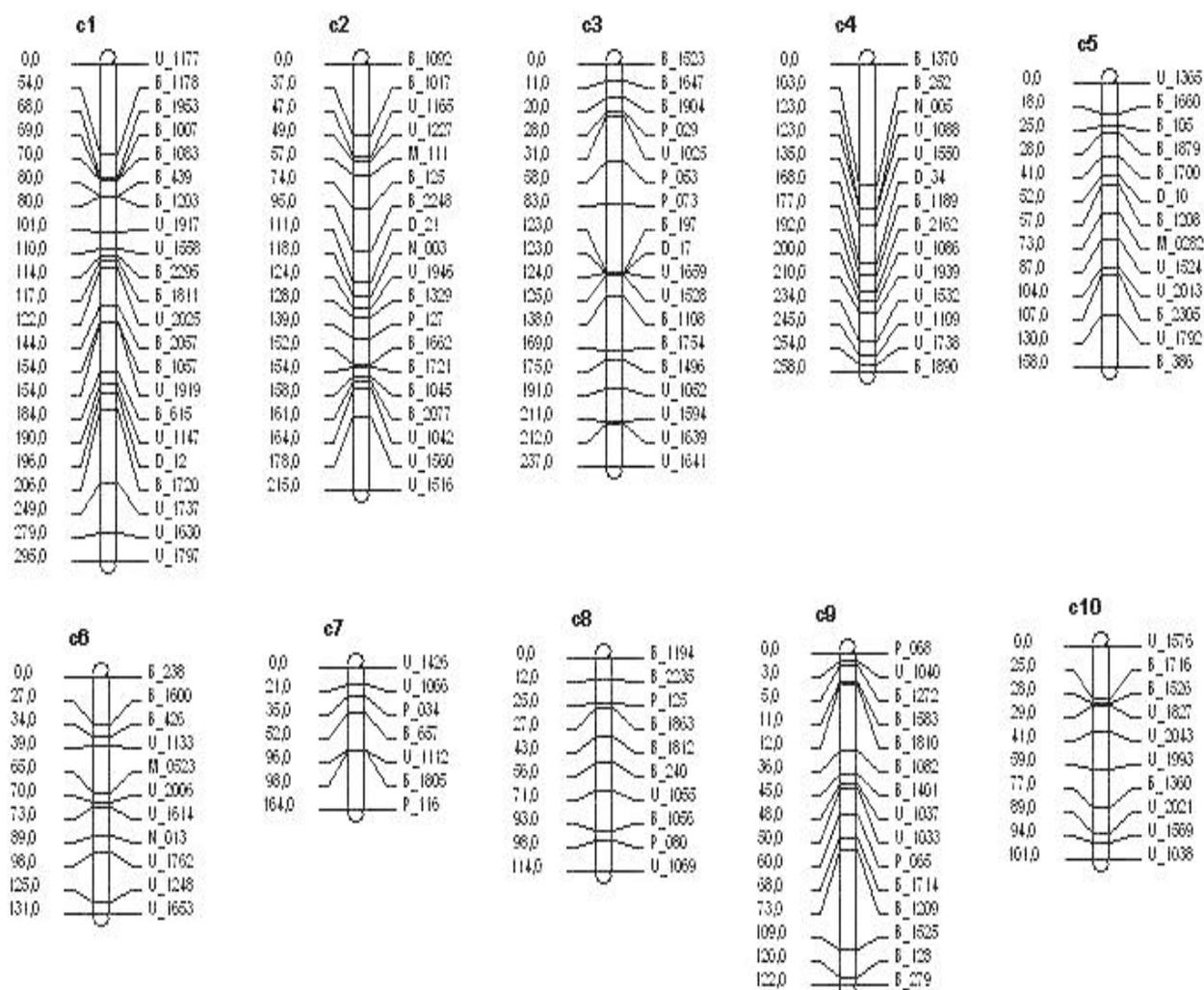


Figura 5 - Mapa genético dos 10 cromossomos de milho com 139 marcas

2.1.10.3 Dados dos caracteres de interesse

O conjunto de dados apresentado na Tabela 3 é um exemplo da constituição genotípica de parte de 171 plantas de milho sob o caráter fenotípico, produção. Cada linha refere-se à informação de uma planta. A primeira coluna mostra o número da planta observada; a segunda coluna trata-se do seu valor fenotípico (produção, em kg/parcela); as demais colunas (M_j , $j=1, \dots, 12$) são as informações genotípicas de doze marcadores moleculares.

Observando a primeira planta de milho, por exemplo, verifica-se que sua produção média foi de 62,5 kg/parcela. O primeiro marcador a ela associado, $M_1 = 2$, é homocigoto dominante (também representado por M_1M_1), o segundo marcador $M_2 = 1$ é heterocigoto

(M_2m_2), enquanto o terceiro $M_3 = 0$ é homozigoto recessivo m_3m_3 , e assim por diante.

Tabela 3 - Dados fenotípicos e genotípicos referentes a produção de milho, em kg/parcela, de 12 marcadores moleculares de parte de 171 plantas de milho, utilizados em um experimento de mapeamento de QTLs

Obs.	Fen.	M_1	M_2	M_3	M_4	M_5	M_6	M_7	M_8	M_9	M_{10}	M_{11}	M_{12}
1	62,5	2	1	0	0	0	0	0	1	1	1	0	0
2	30,0	1	1	1	1	1	2	2	2	2	0	0	1
3	30,0	1	2	2	2	2	1	1	1	1	2	2	2
4	40,0	1	0	0	0	0	0	0	0	0	1	2	2
...
169	55,0	0	1	1	1	1	0	1	1	1	1	1	1
170	30,0	1	0	0	1	1	1	1	1	1	1	1	1
171	57,5	2	2	1	1	0	1	1	1	1	0	0	0

2.1.11 Mapeamento de QTLs

O estudo dos caracteres quantitativos visa identificar e monitorar os locos de interesse ao longo do processo de melhoramento.

O mapeamento de QTLs consiste na sua detecção, localização e estimação de seus efeitos genéticos, tais como: efeito aditivo, efeito de dominância, efeito de epistasia etc. Para isso, é necessário que se possua um mapa de ligação cobrindo todo o genoma e que exista variabilidade para o caráter considerado. A construção do mapa é feita com base no desequilíbrio de ligação existente entre a marca e o QTL. O delineamento experimental considerado para essa finalidade é do tipo F_2 .

O procedimento de mapeamento consiste, portanto, em avaliar os genótipos dos indivíduos da população de interesse. Se houver diferença entre as médias fenotípicas do caráter para cada uma das classes genotípicas de um dado marcador, pode-se inferir que existe um QTL ligado àquela marca (LYNCH; WALSH, 1998).

2.1.12 Métodos Estatísticos para o Mapeamento de QTLs

A associação entre marcador e QTLs pode ser avaliada considerando-se uma, duas ou mais marcas, simultaneamente. Na análise de marcas simples, a distribuição dos valores do caráter é examinada separadamente para cada loco-marcador. Assim, cada teste é realizado independentemente da informação dos outros locos e, deste modo, um cromossomo com m marcas provê m testes de marcas simples (SILVA, 2001).

No mapeamento por intervalo, as análises são realizadas considerando cada par de marcas adjacentes ou flanqueadas, resultando então, em $m - 1$ testes de associação QTL-marcador por cromossomo com m marcas (um para cada intervalo).

Tanto no método de mapeamento por intervalo quanto nas análises de marcas simples, os estimadores associados aos parâmetros genéticos são viesados quando múltiplos QTLs estão ligados à marca. O mapeamento por intervalo composto, é proposto para combinar as técnicas de mapeamento por intervalo e regressão múltipla, no intuito de eliminar as influências de outros QTLs sobre o intervalo considerado.

A técnica de mapeamento por intervalo múltiplo é baseada no modelo genético proposto em 1954 por Cocherham (ZENG, 2002) e no método de máxima verossimilhança para a estimação de parâmetros genéticos. Com esta técnica, pode-se obter um aumento na precisão e no poder da detecção de QTLs, além de permitir a estimação e análise dos efeitos de interação entre QTLs (epistasia).

Em cada método de mapeamento de QTLs, portanto, há diferenças na quantidade e na precisão das informações obtidas.

2.1.12.1 Análise de Marcas Simples

No método de mapeamento utilizando marcas simples, *Single Marker Analysis*, a distribuição dos valores fenotípicos é considerada para cada marcador separadamente, e cada teste é realizado independentemente da informação dos demais locos (ZENG, 1993; 1994; 2001). A comparação entre as médias fenotípicas de cada classe de genótipo marcador pode ser realizada pelo teste t , regressão linear simples, análise de variância ou pelo teste da razão de verossimilhança, LR , do inglês *Likelihood Ratio*. Observada diferença significativa entre as classes de um dado marcador, pode-se inferir sobre a existência de pelo menos um QTL a ele ligado. Esse tipo de análise é bastante simples no que se refere ao aspecto computacional, entretanto apresenta grandes limitações. Não há independência entre os testes de hipóteses

realizados para cada associação marcador-caráter (LIU, 1998), sendo necessário o emprego de alguma correção para que se obtenha o nível de significância conjunto, considerando-se todos os testes.

A maior desvantagem do mapeamento de marcas simples é que o método confunde efeito e posição de QTLs, não havendo distinção entre a ocorrência de um QTL de pequeno efeito situado próximo à marca, ou de um QTL de grande efeito situado mais longe. Com esse tipo de análise, pode-se detectar a existência do QTL associado ao marcador, mas não é possível realizar o mapeamento completo, estimando os efeitos e a posição desses QTLs no genoma. A análise de marca simples pode ser usada, contudo, como um passo inicial ao se realizar o mapeamento empregando-se metodologias mais complexas, uma vez que permite, por exemplo, realizar a eliminação de marcas não informativas, o que aumenta a eficiência do processo (LYNCH; WALSH, 1998; SCHUSTER; CRUZ, 2004). Na abordagem Bayesiana, esse método pode ser considerado como uma análise exploratória para os dados de marcadores.

2.1.12.2 Mapeamento por Intervalo (IM)

O mapeamento por intervalo é um método que, visando a necessidade do desenvolvimento de metodologias mais precisas, foi proposto por Lander e Botstein (1989) com o objetivo de percorrer todo o genoma, fixando posições, na busca por evidências da presença de um QTL. Este método utiliza, como unidade de análise, pares de marcadores adjacentes com genótipos observáveis flanqueando um possível QTL, permitindo, dessa forma, o aumento no poder de detecção, precisão nas estimativas dos efeitos e nas posições dos QTLs. Se for considerado, por exemplo, um mapa elaborado empregando-se m marcas na ordem M_1, M_2, \dots, M_m , haverá um mapa de verossimilhança para cada um dos $m - 1$ intervalos $(M_1 - M_2, M_2 - M_3, \dots, M_{m-1} - M_m)$. Esse mapa utiliza apenas as informações do par de marcadores no intervalo considerado, sendo posteriormente construído o mapa contendo todos os intervalos em conjunto. A análise de ligação entre QTL e marcador é realizada utilizando-se uma função de distância citogenética como a de Haldane, que transforma distâncias em frações de recombinação entre locos e, a partir da informação do genótipo dos marcadores, calcula-se uma determinada estatística para cada um dos $m - 1$ sub-intervalos analisados. No entanto, os efeitos de um QTL detectado em determinado intervalo podem ser afetados por efeitos de outros possíveis QTLs localizados em outras regiões do genoma. Por essa razão os testes para a associação marcador-QTL realizados para cada intervalo não são independentes (ZENG,

1993) e, portanto, o nível de significância utilizado considerando-se todos os intervalos é discutível. Entretanto, existem algumas alternativas para sua obtenção, utilizando, por exemplo, permutações ou estimadores que levam em consideração o número de cromossomos analisados (LYNCH; WALSH, 1998). O mapeamento por intervalo possibilita, portanto, a obtenção de um número maior de informações - posição e efeitos genéticos - bem como possui maior precisão, em relação à análise de marca simples.

Probabilidades Condicionais

O modelo para um único QTL flanqueado por dois marcadores codominantes pode ser aplicado no mapeamento da geração F_2 de um cruzamento entre duas linhagens que carregam alelos diferentes para os três locos sob estudo; dois referentes aos marcadores flanqueadores e um referente ao QTL.

Em uma população produzida pelo cruzamento entre os genitores homocigotos contrastantes M_1QM_2/M_1QM_2 e m_1qm_2/m_1qm_2 , os indivíduos F_1 produzirão gametas com determinada frequência de recombinação. Considerando um QTL Q flanqueado por dois marcadores, M_1 e M_2 , (Figura 6) a frequência dos gametas depende de três parâmetros:

1. A frequência de recombinação entre os marcadores (r_{12});
2. A frequência de recombinação entre o marcador M_1 e o QTL (r_1); e
3. A frequência de recombinação entre o QTL e o marcador M_2 (r_2);

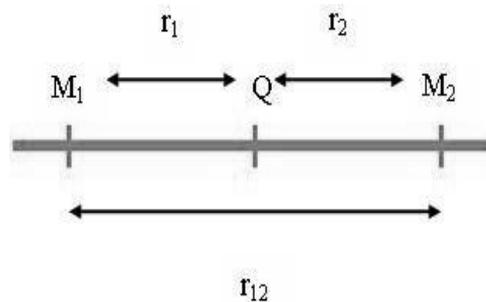


Figura 6 - Marcadores M_1 e M_2 flanqueados ao QTL Q , com suas respectivas frequências de recombinação

Supondo que os genótipos das linhagens parentais L_1 e L_2 cruzadas para um certo conjunto de três locos sejam:

$$L_1 : \frac{M_1QM_2}{M_1QM_2} \times L_2 : \frac{m_1qm_2}{m_1qm_2}$$

A primeira geração, F_1 , apresentará o seguinte genótipo:

$$F_1 : \frac{M_1QM_2}{m_1qm_2}$$

A autofecundação de indivíduos F_1 , produzirá a geração F_2 . Com a ocorrência de permutas gênicas, a geração F_2 poderá apresentar as constituições gaméticas apresentadas na Tabela 4. A probabilidade de cada genótipo da geração F_2 será obtida pelo produto das probabilidades gaméticas de cada combinação de gametas necessária para o genótipo considerado. Por exemplo, a probabilidade de ocorrência do genótipo M_1QM_2/M_1QM_2 é igual ao quadrado da probabilidade de ocorrência do gameta M_1QM_2 e a probabilidade do genótipo M_1QM_2/m_1qm_2 é igual ao produto das probabilidades de ocorrência dos gametas M_1QM_2 e m_1qm_2 .

Tabela 4 - Constituição dos possíveis gametas de indivíduos da geração F_2 obtidas a partir do cruzamento das linhagens parentais L_1 e L_2

Gametas	Probabilidades	Característica
M_1QM_2	$(1 - r_1)(1 - r_2)/2$	Parental
M_1Qm_2	$(1 - r_1)r_2/2$	Recombinante
M_1qM_2	$r_1r_2/2$	Recombinante
M_1qm_2	$r_1(1 - r_2)/2$	Recombinante
m_1QM_2	$r_1r_2/2$	Recombinante
m_1Qm_2	$r_1r_2/2$	Recombinante
m_1qM_2	$(1 - r_1)r_2/2$	Recombinante
m_1qm_2	$(1 - r_1)(1 - r_2)/2$	Parental

Atribuindo valores para r_1 , o valor de r_2 considerando-se a relação (14) e será dado por (eq. 18):

$$r_2 = (r_{12} - r_1)/(1 - 2r_1). \quad (18)$$

Sendo assim, dado o genótipo do marcador M_1M_2/M_1M_2 , tem-se:

$$P \left[\frac{M_1QM_2}{M_1QM_2} \right] = \left[\frac{(1-r_1)(1-r_2)}{2} \right]^2,$$

$$P \left[\frac{M_1QM_2}{M_1qM_2} \right] = 2 \frac{[(1-r_1)(1-r_2)/2]}{(r_1r_2/2)}$$

e

$$P \left[\frac{M_1 q M_2}{M_1 q M_2} \right] = \left(\frac{r_1 r_2}{2} \right)^2.$$

Logo, a probabilidade marginal para os marcadores $M_1 M_2 / M_1 M_2$ será:

$$\begin{aligned}
 P \left[\frac{M_1 M_2}{M_1 M_2} \right] &= P \left[\frac{M_1 Q M_2}{M_1 Q M_2} \right] + P \left[\frac{M_1 q M_2}{M_1 q M_2} \right] + P \left[\frac{M_1 q M_2}{M_1 q M_2} \right] \\
 &= \left[\frac{(1-r_1)(1-r_2)}{2} \right]^2 + 2 \frac{[(1-r_1)(1-r_2)/2]}{(r_1 r_2 / 2)} + \left(\frac{r_1 r_2}{2} \right)^2 \\
 &= \left[\frac{(1-r_1)(1-r_2)}{2} + \frac{r_1 r_2}{2} \right]^2 \\
 &= \frac{1}{4} [(1-r_1)(1-r_2) + r_1 r_2]^2 \\
 &= \frac{1}{4} [1 - r_1 - r_2 + r_1 r_2 + r_1 r_2]^2 \\
 &= \frac{1}{4} [1 - (r_1 + r_2 - 2r_1 r_2)]^2
 \end{aligned} \tag{19}$$

ou seja,

$$P \left[\frac{M_1 M_2}{M_1 M_2} \right] = \frac{(1 - r_{12})^2}{4}. \tag{20}$$

Dessa forma, a probabilidade condicional de Q_k , $k = 1, 2, 3$, dada a ocorrência de $M_j = M_1 M_2 / M_1 M_2$ na população F_2 é dada por:

$$P \left[QQ \left| \frac{M_1 M_2}{M_1 M_2} \right. \right] = \frac{P [M_1 Q M_2 / M_1 Q M_2]}{P [M_1 M_2 / M_1 M_2]} = \frac{[(1-r_1)^2(1-r_2)^2]}{(1-r_{12})^2}, \tag{21}$$

$$P \left[Qq \left| \frac{M_1 M_2}{M_1 M_2} \right. \right] = \frac{P [M_1 Q M_2 / M_1 q M_2]}{P [M_1 M_2 / M_1 M_2]} = \frac{2[(1-r_1)(1-r_2)r_1 r_2]}{(1-r_{12})^2}. \tag{22}$$

e

$$P \left[qq \left| \frac{M_1 M_2}{M_1 M_2} \right. \right] = \frac{P [M_1 q M_2 / M_1 q M_2]}{P [M_1 M_2 / M_1 M_2]} = \frac{r_1 r_2}{(1-r_{12})^2}. \tag{23}$$

sendo $Q_1 = QQ$, $Q_2 = Qq$ e $Q_3 = qq$

Generalizando, as probabilidades condicionais podem ser calculadas por meio de (eq. 24):

$$P[Q_k|M_j] = \frac{P[Q_k, M_j]}{P[M_j]} \quad (24)$$

em que Q_k é o genótipo do QTL e M_j , $j = 1, \dots, m$ o genótipo dos marcadores que estão flanqueando o possível QTL; $P[Q_k, M_j]$ é a probabilidade conjunta de ocorrência do genótipo do QTL e o j -ésimo marcador, M_j e $P[M_j]$ é a probabilidade marginal, isto é, a probabilidade de ocorrência do j -ésimo genótipo marcador.

Na Tabela 5, são apresentadas as probabilidades condicionais associadas aos genótipos do QTL dada a informação dos marcadores flanqueadores M_1 e M_2 , considerando as vinte e sete classes genotípicas de indivíduos da geração F_2 .

Tabela 5 - Probabilidades condicionais associadas aos genótipos de um possível QTL no intervalo entre dois marcadores M_1 e M_2 , em uma geração F_2 , admitindo não interferência

Genótipo Marcador	Genótipo do QTL		
	QQ	Qq	q
$M_1M_1M_2M_2$	$\frac{(1-r_1)^2(1-r_2^2)}{(1-r_{12})^2}$	$\frac{2r_1(1-r_1)r_2(1-r_2)}{(1-r_{12})^2}$	$\frac{r_1^2r_2^2}{(1-r_{12})^2}$
$M_1M_1M_2m_2$	$\frac{2(1-r_1)^2r_2(1-r_2)}{2r_{12}(1-r_{12})}$	$\frac{2r_1(1-r_1)[r_2^2+(1-r_2)^2]}{2r_{12}(1-r_{12})}$	$\frac{2r_1^2r_2(1-r_2)}{2r_{12}(1-r_{12})}$
$M_1M_1m_2m_2$	$\frac{(1-r_1)^2r_2^2}{r_{12}^2}$	$\frac{2r_1(1-r_1)r_2(1-r_2)}{r_{12}^2}$	$\frac{r_1^2(1-r_2)^2}{r_{12}^2}$
$M_1m_1M_2M_2$	$\frac{2r_1(1-r_1)(1-r_2)^2}{2r_{12}(1-r_{12})}$	$\frac{2[r_1^2+(1-r_1)^2]r_2(1-r_2)}{2r_{12}(1-r_{12})}$	$\frac{2r_1(1-r_1)r_2^2}{2r_{12}(1-r_{12})}$
$M_1m_1M_2m_2$	$\frac{4r_1(1-r_1)r_2(1-r_2)}{2[(1-r_{12})^2+r_{12}^2]}$	$\frac{2[r_1^2+(1-r_1)^2][r_2^2+(1-r_2)^2]}{2[(1-r_{12})^2+r_{12}^2]}$	$\frac{4r_1(1-r_1)r_2(1-r_2)}{2[(1-r_{12})^2+r_{12}^2]}$
$M_1m_1m_2m_2$	$\frac{2r_1(1-r_1)r_2^2}{2r_{12}(1-r_{12})}$	$\frac{2[r_1^2+(1-r_1)^2]r_2(1-r_2)}{2r_{12}(1-r_{12})}$	$\frac{2r_1(1-r_1)(1-r_2)^2}{2r_{12}(1-r_{12})}$
$m_1m_1M_2M_2$	$\frac{r_1^2(1-r_2)^2}{r_{12}^2}$	$\frac{2r_1(1-r_1)r_2(1-r_2)}{r_{12}^2}$	$\frac{(1-r_1)^2r_2^2}{r_{12}^2}$
$m_1m_1M_2m_2$	$\frac{2r_1^2r_2(1-r_2)}{2r_{12}(1-r_{12})}$	$\frac{2r_1(1-r_1)[r_2^2+(1-r_2)^2]}{2r_{12}(1-r_{12})}$	$\frac{2(1-r_1)^2r_2(1-r_2)}{2r_{12}(1-r_{12})}$
$m_1m_1m_2m_2$	$\frac{r_1^2r_2^2}{(1-r_{12})^2}$	$\frac{2r_1(1-r_1)r_2(1-r_2)}{(1-r_{12})^2}$	$\frac{(1-r_1)^2(1-r_2)^2}{(1-r_{12})^2}$

Fonte: Schuster et al. (2004)

r_1 é a frequência de recombinação entre o primeiro marcador e o possível QTL; r_2 é a frequência de recombinação entre o possível QTL e o segundo marcador; r_{12} é a frequência de recombinação entre os dois marcadores que flanqueiam o intervalo.

O valor esperado do caráter para os três genótipos possíveis do QTL na geração F_2 são $\mu - a$ (para qq), $\mu + d$ (para Qq) e $\mu + a$ (para QQ) (Figura 7):

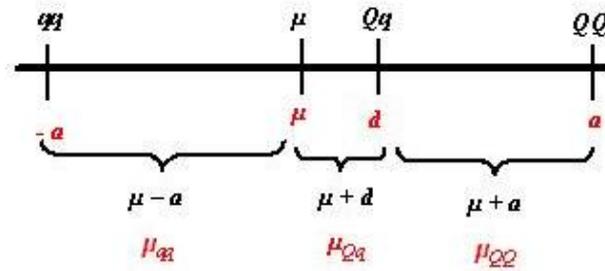


Figura 7 - Escala dos valores genotípicos para um possível QTL, admitindo não interferência

sendo μ a média do caráter em estudo e a e d os desvios devido aos efeitos aditivo e de dominância, respectivamente. A média esperada do caráter quantitativo em termos do suposto QTL, para cada combinação possível do genótipo dos marcadores flanqueadores para indivíduos do cruzamento F_2 , pode ser obtida da seguinte maneira: Partindo do pressuposto que os efeitos ambientais e dos demais QTLs tenham uma distribuição normal, os valores fenotípicos do caráter y dos indivíduos avaliados na geração F_2 serão uma mistura de normais com médias μ_{Q_1} , μ_{Q_2} , μ_{Q_3} e variância comum σ^2 (BEARZOTI, 2000). Desse modo, a distribuição de probabilidade da característica fenotípica do i –ésimo indivíduo dado o genótipo do marcador M_j será dada por (eq. 25):

$$f[y_i|M_j] = \sum_{k=1}^3 N(\mu_{Q_k}, \sigma^2)P[Q_k|M_j], \quad (25)$$

em que M_j : genótipo dos j –ésimo marcador ($M_1M_1M_2M_2$, $M_1M_1M_2m_2$, ..., $m_1m_1m_2m_2$); Q_k : genótipo do QTL Q ($Q_1 = QQ$, $Q_2 = Qq$, $Q_3 = qq$) em uma geração F_2 ; $N(\cdot)$: função densidade de probabilidade normal; $P[Q_k|M_j]$: probabilidade do genótipo do QTL Q_k condicionada às informações dos marcadores flanqueadores (Tabela 5).

Tomando como exemplo um indivíduo i com genótipo $M_j = M_1M_1M_2M_2$, a densidade de probabilidade da característica fenotípica será (eq. 26):

$$f[y_i|M_1M_1M_2M_2] = N(\mu_{QQ}, \sigma^2)P\left[QQ\left|\frac{M_1M_2}{M_1M_2}\right.\right] + N(\mu_{Qq}, \sigma^2)P\left[Qq\left|\frac{M_1M_2}{M_1M_2}\right.\right] + N(\mu_{qq}, \sigma^2)P\left[qq\left|\frac{M_1M_2}{M_1M_2}\right.\right]. \quad (26)$$

O valor esperado da característica fenotípica para o i –ésimo indivíduo da geração F_2 será:

$$\begin{aligned}
E[Y_i|M_1M_1M_2M_2] &= (\mu_{QQ})P \left[QQ \left| \frac{M_1M_2}{M_1M_2} \right. \right] + (\mu_{Qq})P \left[Qq \left| \frac{M_1M_2}{M_1M_2} \right. \right] + (\mu_{qq})P \left[qq \left| \frac{M_1M_2}{M_1M_2} \right. \right] \\
&= (\mu + a)P \left[QQ \left| \frac{M_1M_2}{M_1M_2} \right. \right] + (\mu + d)P \left[Qq \left| \frac{M_1M_2}{M_1M_2} \right. \right] + (\mu - a)P \left[qq \left| \frac{M_1M_2}{M_1M_2} \right. \right] \\
&= (\mu + a) \left\{ P \left[QQ \left| \frac{M_1M_2}{M_1M_2} \right. \right] - P \left[qq \left| \frac{M_1M_2}{M_1M_2} \right. \right] \right\} + (\mu + d)P \left[Qq \left| \frac{M_1M_2}{M_1M_2} \right. \right]
\end{aligned} \tag{27}$$

das expressões 21, 22 e 23, tem-se:

$$E[Y_i|M_1M_1M_2M_2] = \mu + a \left\{ \frac{(1 - r_1)^2(1 - r_2^2)}{(1 - r_{12})^2} - \frac{r_1^2 r_2^2}{(1 - r_{12})^2} \right\} + d \frac{2r_1(1 - r_1)r_2(1 - r_2)}{(1 - r_{12})^2}, \tag{28}$$

ou ainda,

$$E[Y|M_1M_1M_2M_2] = \mu + aX_a + dX_d. \tag{29}$$

Considerando-se que a fração de recombinação entre os marcadores M_1 e M_2 (r_{12}) é conhecida (informação do mapa de ligação), a estratégia do mapeamento de QTL por intervalo consiste em atribuir valores para r_1 . A seguir, estimam-se os valores para os parâmetros μ , a , d .

2.1.12.3 Mapeamento por Intervalo Composto (CIM)

No mapeamento de QTLs, o mapeamento por intervalos simples que avalia um intervalo de cada vez, sem considerar o restante da informação do genoma, apresenta problemas de detecção, especialmente quando houver outros QTLs de grande efeito no mesmo ou em outros grupos de ligação. Desse modo, não há como determinar se o efeito significativo observado é devido a apenas um QTL ou a vários QTLs ligados. Além disso, quando múltiplos QTLs estão presentes, um problema comum é a observação de falsos positivos, ou seja, são detectados QTLs inexistentes, ou “QTLs fantasmas”.

Com o objetivo de eliminar os efeitos dos QTLs inexistentes, foi proposta, de forma independente por Jansen (1993, 1995) e Zeng (1993, 1994), a utilização de um modelo de regressão múltipla, constituindo, então, o mapeamento por intervalo composto (*Composite Interval Mapping* - CIM). Esse método combina mapeamento por intervalo e regressão linear

múltipla, foi construído com o propósito de evitar que um QTL presente em qualquer posição do genoma interfira no mapeamento do QTL no intervalo considerado. Isso é feito incluindo no modelo de regressão múltipla a informação de marcadores adjacentes ao intervalo considerado e de marcadores presentes em outros grupos de ligação, sob a forma de cofatores (covariáveis). Considerando o segmento cromossômico esquematizado na Figura 8, suponha que o intervalo analisado pelo método CIM esteja entre as marcas M_j e $M_{(j+1)}$. Adicionando como cofatores os marcadores adjacentes $M_{(j-1)}$ e $M_{(j+2)}$, estes impedem que os efeitos do QTL Q_1 ligado à esquerda da marca $M_{(j-1)}$ e do QTL Q_3 à direita de $M_{(j+2)}$ interfiram na detecção do QTL Q_2 existente no intervalo de interesse ($j, j + 1$) (SILVA, 2001):

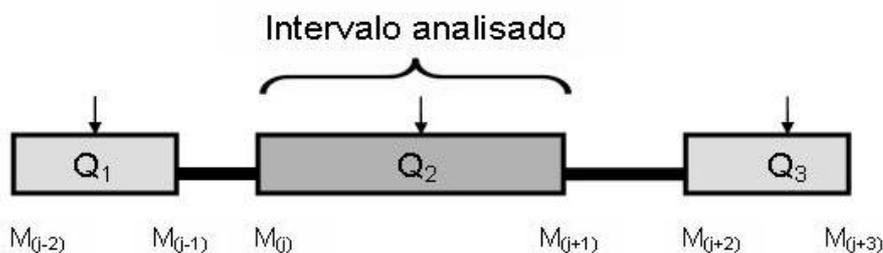


Figura 8 - Esquema de um cromossomo mostrando três QTLs (Q_1, Q_2, Q_3) flanqueados por diferentes marcas $M_{(j-2)}; M_{(j-1)}$, $M_{(j)}; M_{(j+1)}$ e $M_{(j+2)}; M_{(j+3)}$. O intervalo de interesse para mapear Q_2 é ($j, j + 1$)

Esse método, porém não impede que os QTLs localizados nos intervalos ($j - 1, j$) e ($j + 1, j + 2$) interfiram no mapeamento ocorrido no intervalo ($j, j + 1$). Com o uso de mapas de ligação saturados (pequenas distâncias entre marcas), a probabilidade de ocorrência de QTLs adjacentes torna-se baixa, e com isso há um substancial incremento na precisão das estimativas de sua posição e efeitos genéticos (LYNCH; WALSH, 1998). No entanto, uma fragilidade do procedimento CIM está em ignorar o efeito de epistasia gênica, ou seja, o efeito de interação entre alelos de diferentes locos é ignorado (ZENG, 1994).

Seleção dos cofatores não adjacentes

Um fator de fundamental importância no mapeamento por intervalo composto é a seleção dos cofatores (covariáveis), que correspondem aos marcadores que possuem maior associação com o caráter (ZENG, 1994) e estes não deverão exceder o número de $2\sqrt{n}$, sendo n o número de observações. Por esse critério, a quantidade de marcadores utilizados no presente trabalho seria de, no máximo, $2\sqrt{256} = 32$. Para selecioná-los, vários métodos podem ser utilizados, como o “*forward*”, “*backward*” e o “*stepwise*”.

O método seqüencial de seleção “*forward*” inicia-se com a suposição de que ne-

nhum marcador está ligado com o caráter em estudo. Por essa razão, a busca pelo melhor subconjunto de cofatores inicia-se pelo modelo nulo. A primeira covariável a ser indicada para entrar no modelo será de acordo com algum critério adotado, como por exemplo, aquela que apresentar maior coeficiente de correlação simples com o caráter em estudo e a seguir covariáveis, uma a uma, serão indicadas se apresentarem a maior estatística $F_{parcial}$. O procedimento de busca é interrompido quando não houver mais covariáveis com estatística $F_{parcial}$ maior do que F_{IN} , determinada que é o percentil da tabela F com 1 e $n - m_c - 1$ graus de liberdade. Nesse caso, o modelo terá n observações e m_c marcadores selecionados como cofatores.

O método seqüencial de eliminação “*backward*” inicia-se com a suposição de que todos os marcadores estão associados ao caráter. Por essa razão, a busca pelo melhor conjunto de cofatores inicia-se pelo modelo maximal (modelo com todos os cofatores). A primeira covariável a ser indicada para sair do modelo poderá ser aquela que apresentar o menor valor da estatística $F_{parcial}$, de acordo com algum critério de parada.

O “*stepwise*” trata-se de uma combinação entre os dois métodos anteriores, em que a busca pelo melhor subconjunto de cofatores parte do modelo nulo, pois como o método “*forward*”, a suposição inicial é a de que nenhum dos marcadores estão associados ao caráter. Os marcadores são adicionados ao modelo, um a um, por ordem de associação com o caráter, que pode ser determinada pelo coeficiente de correlação, correlação parcial ou coeficiente de regressão linear simples, sendo que a sua permanência depende da significância de seu efeito. A cada inclusão de um novo marcador são observadas as significâncias dos efeitos dos demais, presentes no modelo, eliminando-se aqueles menos informativos, obtendo-se, assim, o conjunto de cofatores que melhor explica a variação do caráter de acordo com o modelo de regressão múltipla.

Assim, o mapeamento por intervalo composto constitui-se na metodologia mais informativa e precisa, dentre as citadas anteriormente.

Estimação dos parâmetros utilizando o método de Mínimos Quadrados

Em uma amostra de n indivíduos avaliados em relação a uma característica quantitativa e genotipados para m marcadores, o modelo de uma regressão linear múltipla para o mapeamento por intervalo composto é dado por (eq. 30):

$$Y_i|(M_j, M_{j+1}) = \mu + \underbrace{aX_{ai} + dX_{di}}_{\text{Intervalo}} + \underbrace{\sum_{c \neq j, j+1}^{m_c} a_c X_{ac} + \sum_{c \neq j, j+1}^{m_c} d_c X_{dc}}_{\text{Cofatores}} + \varepsilon_i \quad (30)$$

sendo Y_i : média geral da característica quantitativa avaliada, $i = 1, \dots, n$; M_j, M_{j+1} : genótipo do i – *ésimo* indivíduo para o intervalo considerado, $j = 1, \dots, m$; a : coeficiente de regressão parcial dos efeitos aditivos do QTL; d : coeficiente de regressão parcial dos efeitos de dominância; a_c e d_c : coeficientes de regressão parciais dos efeitos aditivo e de dominância para o caráter em questão na marca c (M_c) a ser utilizada como covariável obtida pelo processo de seleção de cofatores através de uma regressão linear múltipla; β : vetor de parâmetros, considerando-se, por exemplo, apenas uma marca M_c , $\beta^T = [\mu, a, d, a_c, d_c]$; $\varepsilon_i \sim N(0, \sigma^2)$: erro aleatório manifestado na característica do i – *ésimo* indivíduo; X_{ai} e X_{di} : colunas de \mathbf{X} , contendo os valores preditos para os efeitos aditivo e de dominância do QTL, calculados por meio das informações genótípicas dos marcadores flanqueadores do suposto QTL e por meio das probabilidades condicionais associadas; X_{ac} : variável indicadora para o efeito aditivo do c – *ésimo* código marcador fora do intervalo considerado (M_c) para o i – *ésimo* indivíduo ($MM = 1, Mm = 0$, e $mm = -1$); X_{dc} : variável indicadora para o efeito de dominância do c – *ésimo* código marcador fora do intervalo para o i – *ésimo* indivíduo ($MM = 0, Mm = 1$, e $mm = 0$); m_c : número total de marcadores utilizados como cofatores.

Como se observa, a primeira parte do somatório na expressão (30) refere-se aos efeitos do QTL dentro do intervalo analisado, enquanto a segunda refere-se aos possíveis efeitos dos demais QTLs fora do mesmo. A matriz de delineamento para o exemplo de apenas um cofator M_c , é dada por:

$$\mathbf{X}_{(n \times 5)} = \begin{bmatrix} 1 & \left(P \left[QQ \left| \frac{M_1 M_2}{M_1 M_2} \right. \right] - P \left[qq \left| \frac{M_1 M_2}{M_1 M_2} \right. \right] \right) & P \left[Qq \left| \frac{M_1 M_2}{M_1 M_2} \right. \right] & 1 & 0 \\ 1 & \left(P \left[QQ \left| \frac{M_1 m_2}{M_1 m_2} \right. \right] - P \left[qq \left| \frac{M_1 m_2}{M_1 m_2} \right. \right] \right) & P \left[Qq \left| \frac{M_1 m_2}{M_1 m_2} \right. \right] & -1 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 1 & \left(P \left[QQ \left| \frac{m_1 M_2}{m_1 M_2} \right. \right] - P \left[qq \left| \frac{m_1 M_2}{m_1 M_2} \right. \right] \right) & P \left[Qq \left| \frac{m_1 M_2}{m_1 M_2} \right. \right] & 0 & 1 \end{bmatrix}.$$

em que a primeira coluna de \mathbf{X} é o vetor associado à média μ ; a segunda e terceira colunas referem-se às probabilidades associadas aos efeitos aditivos a e de dominância d ; a quarta e quinta colunas da matriz \mathbf{X} , estão associadas aos códigos do genótipo do cofator M_c fora do

intervalo analisado para os efeitos de aditividade e dominância, a_c e d_c respectivamente.

Ajustando-se o modelo de regressão com médias dos indivíduos para cada classe genotípica dos marcadores flanqueados ao QTL, o modelo será $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, e a matriz $\mathbf{X}_{(n \times 5)}$ de delineamento terá as informações das probabilidades dos QTLs condicionados pelas marcas a eles adjacentes e os genótipos dos marcadores fora do intervalo estudado.

A estimativa dos parâmetros genéticos, pelo método de mínimos quadrados para cada valor de r_1 no sub-intervalo r_{12} conhecido é

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}. \quad (31)$$

Estimação dos parâmetros utilizando os métodos de Máxima Verossimilhança

Visto que os valores fenotípicos do caráter \mathbf{y} dos indivíduos avaliados podem seguir uma mistura de Normais, devido aos efeitos ambientais e das três diferentes classes de genótipos do QTL em uma geração F_2 (Figura 9).

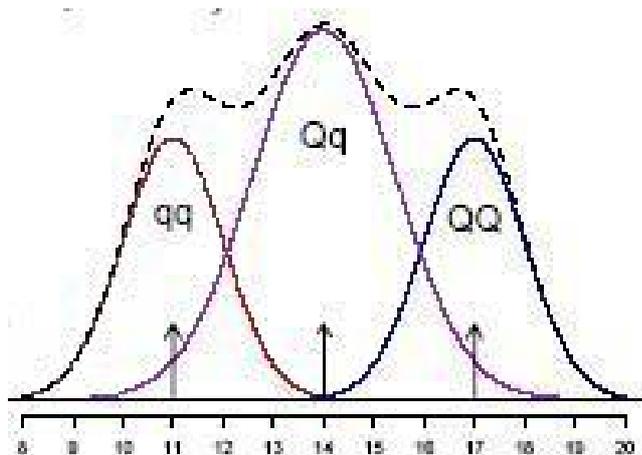


Figura 9 - Esquema de um conjunto de dados que segue um modelo de mistura de densidades Normais

A função de verossimilhança, considerando os genótipos marcadores $M_1M_1M_2M_2$ sob independência, é o produto das contribuições individuais de verossimilhanças:

$$f[y_i | M_1M_1M_2M_2] = \sum_{k=1}^3 N(\mu_Q, \sigma^2) P[Q_k | M_j]$$

E assim a função de verossimilhança é dada por:

$$L(r_1, \mu, a, d, \sigma^2; y_i) = \prod_{i=1}^n \frac{1}{\sqrt{(2\pi\sigma^2)}} \sum_{k=1}^3 P[Q_k | M_j] e^{-\frac{(y_i - \mu - a_k - d_k)^2}{2\sigma^2}}. \quad (32)$$

Para o caso de mistura de Normais, os estimadores de máxima verossimilhança para μ , a e d não têm solução explícita e devem ser obtidos por métodos numéricos. Para a obtenção dos estimadores de máxima verossimilhança, Zeng (1994) sugere a utilização do algoritmo ECM (“*Expectation Conditional Maximization*”), que consiste em fixar a posição de um eventual QTL dentro do intervalo e, em seguida, estimar os demais parâmetros.

Detecção de QTLs

O teste estatístico para detectar a presença de um QTL é realizado movendo-se a posição do suposto QTL ao longo do cromossomo. Para uma dada relação entre distância e fração de recombinação em relação a um dos marcadores (r_i). O teste utilizado para verificar se existe associação entre QTL e marcador é a razão de verossimilhança.

A estatística do teste da razão de verossimilhanças é dada por (33)

$$LR = 2ln \left[\frac{\text{máx. } L(H_a)}{\text{máx. } L(H_0)} \right] = -2ln \left[\frac{\text{máx. } L(H_0)}{\text{máx. } L(H_a)} \right]. \quad (33)$$

sendo $L(H_0)$ a função de verossimilhança sob a hipótese nula e $L(H_a)$ a função de verossimilhança sob a hipótese alternativa.

Na análise de Regressão quando os erros são independentes e identicamente distribuídos, o teste da razão de verossimilhanças pode ser expresso em função do total de observações (n) e das Somas de Quadrados de Resíduos para os modelos completo (SQ_ϵ Modelo Completo) e reduzido (SQ_ϵ Modelo Reduzido). Para o exemplo utilizando apenas um cofator, tem-se (34):

$$LR = n \left[\frac{SQ_\epsilon \text{Modelo Reduzido } (\mu, a, d)}{SQ_\epsilon \text{Modelo Completo } (\mu, a, d, a_c, d_c)} \right]. \quad (34)$$

Admitindo-se não interferência, LR é estimado para cada valor de r_1 no intervalo.

A estatística do teste da razão de verossimilhanças definida pela expressão (33) tem distribuição assintótica de qui-quadrado com d graus de liberdade e nível de significância para cada teste $\alpha' = \alpha/(m - 1)$, ou seja, $LR \sim \chi_{\alpha', d}^2$, sendo α o nível de significância global adotado e $m - 1$ o número de intervalos considerado e, portanto, o número de testes realizados. O número de graus de liberdade apropriado para o teste deve ser definido como a diferença

entre as dimensões do espaço paramétrico completo sob H_a e do do espaço paramétrico restrito sob H_0 (LYNCH; WALSH, 1998; BEARZOTI, 2000).

No mapeamento de QTLs o teste da razão de verossimilhanças também pode ser expresso através do *LODscore* (“*Likelihood of odds score*”) utilizado para declarar se um QTL está presente ou não em determinada posição do genoma, além de facilitar a compreensão da magnitude do teste pelo uso da escala logarítmica. A diferença entre *LODscore* e a estatística *LR* é a base do logaritmo usada no teste. O teste *LR* utiliza o logaritmo neperiano e o *LODscore* utiliza o logaritmo na base 10. Como visto, a estatística *LR* é interpretada como a probabilidade de ocorrência dos dados, sob a hipótese nula, a qual é testada usando a distribuição teórica qui-quadrado. O *LODscore* não necessita de uma distribuição teórica para a sua interpretação. Por exemplo, *LODscore* igual a três indica que a hipótese alternativa, referente à ocorrência de um QTL, é $10^3 = 1000$ vezes mais provável de ocorrer do que a hipótese nula da não ocorrência. Isto significa que, quando a estatística *LODscore* apresentar um valor igual a três para determinada posição, haverá indicação de que existe um QTL naquela região cromossômica com a probabilidade de um resultado falso positivo igual a 0,001. Em trabalhos de mapeamento de QTLs na cultura de milho, os limites críticos empregados possuem valores de *LODscore* entre dois e três.

A relação entre os valores do *LR* e do *LODscore* (eq. 35) é dada por (LANDER; BOTSTEIN, 1999; LYNCH; WALSH, 1998):

$$LOD = \frac{LR}{2 \ln 10} \simeq \frac{LR}{4,61} \quad (35)$$

Adotando-se, por exemplo, um valor para o limite crítico $LR = 11,5$, da expressão (35) tem-se o valor aproximado de 2,5 para o *LOD*.

Determinar o limite crítico ou o “*threshold*” a ser utilizado, é uma questão de muita importância para o mapeamento de QTLs. A posição do QTL no mapa é estimada usando a estatística *LR* como uma função de posição de mapa do suposto QTL. As probabilidades de detecção e estimativa da posição de um QTL ligado estão acopladas. Se a razão de verossimilhanças atingir o valor máximo dentre os que excederam o limiar crítico para aquele cromossomo, isso fornece evidência de QTL no intervalo, sendo a posição estimada pelo pico da posição mais verossímil. Se o pico não exceder o limiar crítico, não existem evidências para aceitar a presença de QTLs na região.

Os métodos de marcas simples e mapeamento por intervalo simples assumem apenas um QTL ligado a um marcador de interesse, não podendo detectar a presença de múltiplos QTLs e, portanto, discernir se um efeito significativo em vários intervalos ligados por marcadores é devido a um QTL comum ou vários QTLs ligados.

O mapeamento por intervalo composto, ao incluir no modelo a informação dos demais QTLs que afetam o carácter em estudo como cofatores (covariáveis), oferecem maior poder de detecção e discriminação de QTL do que as metodologias anteriores (análise de marcas simples e mapeamento por intervalo simples).

O mapeamento por intervalo composto realiza $m - 1$ testes em um mapa genético composto de m marcas. No entanto, o fato haver um nível α de significância associado em cada teste não implica que o nível de significância global para a família de comparações também seja igual a α . Por essa razão o valor de α deve ser corrigido para evitar a detecção de QTLs inexistentes. Uma alternativa comum está em utilizar a correção de Bonferroni que estabelece um nível de significância α para todo o conjunto de $(m - 1)$ testes e $\alpha' = \alpha / (m - 1)$ para cada um deles, pois se o nível desejado é α , então $\alpha = (m - 1)\alpha'$ (BEARZOTI, 2000; BARBIN, 2003).

2.1.12.4 Mapeamento por Intervalo Bayesiano (BIM)

Em uma população F_2 , considere $\mathbf{y}^T = (y_1, \dots, y_n)$ um vetor de valores fenotípicos do carácter avaliado, observado em n indivíduos e P uma variável aleatória, que denota o número de QTLs. O mapeamento por intervalo Bayesiano (BIM, “*Bayesian Interval Mapping*”) considera o seguintes modelo estatístico (SATAGOPAN, 1996):

$$y_i = \mu + \sum_{j=1}^p a_j Q_{ij} + \sum_{j=1}^p d_j (1 - |Q_{ij}|) + e_i \quad (36)$$

sendo y_i o valor fenotípico do i -ésimo indivíduo ($i = 1, \dots, n$), μ a média geral do carácter, $e_i \sim N(0, \sigma^2)$ o erro aleatório, $a = \{a_j\}_{j=1}^p$ e $d = \{d_j\}_{j=1}^p$ denotam os efeitos de aditividade e dominância para o QTL j ; $Q_i = \{Q_{ij}\}_{j=1}^p$ denota o vetor de genótipos dos P QTLs para o indivíduo i , em que $Q_{ij} = -1$ se $Q_{ij} = qq$, $Q_{ij} = 0$ se $Q_{ij} = Qq$ e $Q_{ij} = 1$ se $Q_{ij} = QQ$.

Uma notação equivalente é dada por:

$$\mathbf{y} = \mathbf{X}^{(p)} \boldsymbol{\Theta}^{(p)} + \boldsymbol{\varepsilon} \quad (37)$$

em que \mathbf{y} é o vetor de valores fenotípicos de n indivíduos; $\boldsymbol{\theta}^{(p)} = (\mu, \sigma^2, a_1, d_1, a_2, d_2, \dots, a_p, d_p)^T$ é o vetor de parâmetros desconhecidos, sendo a_j e d_j associados aos efeitos aditivo e dominância do QTL, com $2p+1$ colunas (se o número de QTLs for considerado aleatório, conseqüentemente o comprimento do vetor de parâmetros, $\boldsymbol{\theta}^{(p)}$, também o será); $\boldsymbol{\varepsilon}$ é o vetor de erros, normalmente distribuídos, $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \mathbf{I}\sigma^2)$.

A matriz de delineamento $\mathbf{X}^{(p)}$ com n linhas e $2p+1$ colunas (GAFFNEY, 2001) que incluirá todas as informações genotípicas do QTL é dada por:

$$\mathbf{X}^{(p)} = \left[\mathbf{1} \quad \mathbf{Q}_{i1} \quad (1 - |\mathbf{Q}_{i1}|) \quad \mathbf{Q}_{i2} \quad (1 - |\mathbf{Q}_{i2}|) \quad \dots \quad \mathbf{Q}_{ip} \quad (1 - |\mathbf{Q}_{ip}|) \right]$$

Assumindo que $e_i \sim N(0, \sigma^2)$, os parâmetros genéticos de interesse são o número de QTLs (P), sua localização $\lambda = \{\lambda_j\}_{j=1}^p$ e $\boldsymbol{\theta}^{(p)} = (\mu, \sigma^2, a_1, d_1, a_2, d_2, \dots, a_p, d_p)^T$, sendo λ_j a posição do j -ésimo QTL no genoma. O modelo também permite que efeitos de epistasia entre QTLs sejam incorporados.

Na prática, são observadas apenas a característica fenotípica y_i e o genótipo dos marcadores M_i . Supondo que o mapa de ligação foi construído, a distância genética $D = \{D_k\}_{k=1}^{(m-1)}$ entre o marcador 1 e o marcador k pertencentes ao mesmo grupo de ligação pode ser observada (Figura 10):

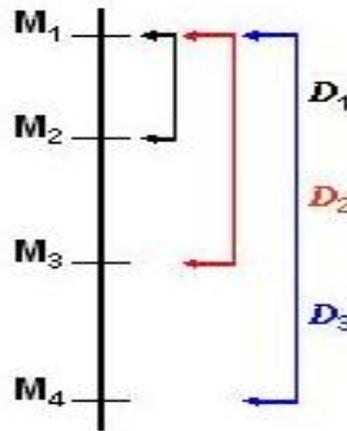


Figura 10 - Esquema de um cromossomo mostrando quatro marcadores e as distâncias genéticas entre o primeiro marcador e o marcador k

O genótipo do QTL Q_i é desconhecido. Entretanto, dadas as informações fornecidas pelo mapa de ligação sobre a localização do QTL (λ), os genótipos dos marcadores $M = \{M_{ik}\}_{k=1}^m$, $i = 1, \dots, n$ e a distância D entre os marcadores, pode-se modelar a distribuição

de probabilidade para o genótipo do QTL em termos de fração de recombinação entre o QTL e marcador.

Suponha que o j – *ésimo* QTL está flanqueado pelos marcadores k_j e $k_{(j+1)}$ e que nenhum outro QTL pertença a esse intervalo (essa é uma suposição razoável se o mapa for saturado). Sob a hipótese de independência dos eventos de recombinação, o genótipo desse QTL será condicionado às informações dos genótipos dos marcadores flanqueadores k_j e $k_{(j+1)}$ e das probabilidades condicionais associadas. Dessa maneira, considerando o modelo apresentado na expressão (36), a função de probabilidade do genótipo do QTL Q_i para o i – *ésimo* indivíduo é dada por:

$$f(Q_i|\lambda, M_i, D) = \prod_{i=1}^p f(Q_{ij}|\lambda_j, M_i, D)$$

Sob a suposição de segregações independentes, admitida pela função de Haldane, a probabilidade condicional do genótipo do QTL flanqueado pelas marcas k_j e k_{j+1} é independente para os genótipos de marcadores pertencentes a diferentes intervalos. A função densidade de probabilidade do genótipo do QTL para o i – *ésimo* indivíduo será:

$$\prod_{j=1}^p f(Q_{ij}|\lambda_j, M_{ik_j}, M_{ik_{(j+1)}}, D_{k_j}, D_{k_{(j+1)}}). \quad (38)$$

sendo $D_{k_j} \leq \lambda_j < D_{k_{(j+1)}}$.

A função de verossimilhança da localização λ do QTL e do parâmetro θ para o i – *ésimo* indivíduo pode ser expressa como (eq. 39):

$$L(\lambda, \theta|y_i, M_i, D) = \sum_{q_i} f(y_i, Q_i = q_i|\lambda, \theta, M_i, D),$$

ou ainda,

$$L(\lambda, \theta|y_i, M_i, D) = \sum_{q_i} f(y_i|Q_i = q_i, \theta)f(Q_i = q_i|\lambda, M_i, D), \quad (39)$$

com a soma sobre o conjunto de todos os possíveis genótipos do QTL para o i – *ésimo* indivíduo, $q_i \in \{-1, 0, 1\}^p$. Considerando-se o vetor de n observações independentes $\mathbf{y} = (y_1, \dots, y_n)^T$ de uma amostra aleatória, teremos a função de verossimilhança (eq. 40):

$$L(\lambda, \theta|\mathbf{y}) = \prod_{i=1}^n \sum_{q_i} f(y_i|Q_i = q_i, \theta)\pi(Q_i = q_i|\lambda, M_i, D) \quad (40)$$

o qual é um modelo finito de mistura de densidades, o que torna a inferência sobre os parâmetros de interesse ligeiramente complexa.

O objetivo é fazer inferências sobre a posição do QTL (λ) e sobre seus efeitos $\boldsymbol{\theta}^{(p)} = (\mu, \sigma^2, a_1, d_1, a_2, d_2, \dots, a_p, d_p)^T$. Na abordagem Bayesiana as inferências sobre os parâmetros genéticos de interesse estão baseadas na distribuição conjunta *a posteriori*, dada pela expressão (eq. 41):

$$\pi(\lambda, \boldsymbol{\theta}, Q|\mathbf{y}) \propto L(\lambda, \boldsymbol{\theta}|\mathbf{y})\pi(\lambda, \boldsymbol{\theta}) \quad (41)$$

sendo $\pi(\boldsymbol{\lambda}, \boldsymbol{\theta})$ a densidade *a priori* para os parâmetros genéticos.

Assumindo independência entre os parâmetros, tem-se (eq. 42):

$$\pi(\boldsymbol{\lambda}, \boldsymbol{\theta}) = \pi(\boldsymbol{\lambda})\pi(\mu)\pi(\sigma^2) \prod_{j=1}^p [\pi(a_j)\pi(d_j)] \quad (42)$$

De acordo com Satagopan et al. (1996), uma escolha natural para *a priori* de λ quando não existe informação sobre a localização dos QTLs é a distribuição uniforme. Para μ, a_j, d_j e σ^2 admitem-se *prioris* conjugadas.

Os métodos de amostragem via cadeias de Markov são comumente usados para calcular integrais complexas envolvendo verossimilhanças, na obtenção dos resumos de densidades *a posteriori*, $\pi(\boldsymbol{\theta}|\mathbf{y})$, a partir de distribuições *a priori* especificadas para os efeitos de aditividade, dominância, de localização no genoma e do erro. A distribuição estacionária é a densidade *a posteriori* de interesse.

A suposição de que o número de QTL é desconhecido, torna a obtenção da amostra da distribuição conjunta *a posteriori* ligeiramente mais complicada. Uma solução para esse problema está na utilização de algoritmos trans-dimensionais projetados para construir cadeias de Markov onde a dimensão do espaço de estados muda ao longo das iterações, tais como algoritmos MCMC com saltos reversíveis, “*Reversible Jump MCMC*” (RJMC) (GREEN, 1995). Esses métodos são particularmente úteis no contexto de seleção Bayesiana de modelos, para estimação pontual e intervalar dos efeitos aditivo e de dominância dos QTLs e sua localização. Foram projetados para construir uma cadeia de Markov em que a dimensão do espaço de estados muda ao longo das iterações, devido ao acréscimo ou decréscimo do número de QTLs no modelo. Esse método, considerado uma generalização do algoritmo Metropolis-Hastings, supõe que o estado atual da cadeia de Markov tenha $P = p$ QTLs, com probabilidade

$pr_{p,p-1}$ para remover um QTL (processo de morte, “*death step*”) e, $pr_{p,p+1}$ para o acréscimo de um QTL no modelo (processo de nascimento, “*birth step*”). A proposta de nascimento não garante a aceitação de um QTL, ou seja, se um QTL foi acrescentado durante o processo isto não significa que ele será incorporado no modelo. (GREEN, 1995). A atualização dos parâmetros de interesse durante o processo de nascimento e morte é feita com probabilidade $1 - pr_{p,p-1} - pr_{p,p+1}$.

Para estimar o número de QTLs no genoma que afetam determinado caráter, uma alternativa é utilizar o Fator de Bayes como critério Bayesiano para seleção de modelos. Diferentes modelos devem ser comparados dois a dois através do Fator de Bayes calculado, (eq. 6); por exemplo, o modelo com 0 QTL contra o modelo com 1 QTL ($M_0 \times M_1$), ..., o modelo com $(p - 1)$ QTLs contra o modelo com p QTLs ($M_{(p-1)} \times M_p$).

2.1.12.5 Especificação das *Prioris*

A especificação das densidades *a priori* é essencial na análise Bayesiana, pois consiste em explicitar a quantidade e a eficiência do conhecimento prévio sobre os dados observados. Em mapeamento de QTLs, serão especificadas distribuições *a priori* para os parâmetros de interesse associados ao número e à posição dos QTLs, bem como aos seus efeitos de aditividade e de dominância.

Número total de QTLs

Freqüentemente, assume-se uma distribuição *a priori* Poisson com média k para a variável aleatória P associada ao número total de QTLs no modelo (SILLANPÄÄ; ARJAS, 1998). Entretanto, a dificuldade está em estabelecer o valor para k , pois este exerce grande influência sobre o valor *a posteriori* para o total de QTLs. Sillanpää e Arjas (1998) sugerem uma distribuição Poisson com média dois ou no máximo três QTLs por cromossomo. O módulo *Bmapqtl* do *software QTLCartographer* (WANG; BASTEN; ZENG, 2005) recomenda *priori* Poisson com média três, ou seja,

$$P \sim Po(3).$$

Um genoma contendo p QTLs e c cromossomos dado por $D(p, c)$, dispõe de uma grande variedade de modelos com p QTLs distribuídos em c cromossomos. O número de modelos $D(p, c)$, $\#D(p, c)$, será:

$$\#D(p, c) = \binom{p + c - 1}{c - 1}.$$

A distribuição *a priori* para o número de QTLs em cada cromossomo é obtida conjuntamente para o vetor $\mathbf{p}^T = (p_1, \dots, p_c)$ sendo c o número de cromossomos no genoma e p_i o número de QTLs no cromossomo i , em que $\sum_{i=1}^c p_i = p$.

Suponha que cada QTL tenha a mesma probabilidade de estar localizado em todo o genoma de comprimento L , e que o cromossomo i tenha comprimento L_i . Então, $\mathbf{p}^T|p$ terá distribuição multinomial

$$\pi(\mathbf{p}^T) = \pi(\mathbf{p}^T, p) = \pi(\mathbf{p}^T|p)\pi(p),$$

em que $\pi(p)$ é a distribuição *a priori* para o número de QTLs (usualmente Poisson com média $k = 3$); $\pi(\mathbf{p}^T)$ é a distribuição *a posteriori* para o número de QTLs em cada um dos c cromossomos, dada a informação de que todo o genoma contenha o total de p QTLs.

Localização do QTL no cromossomo

A posição do j -ésimo QTL, λ_j , é especificada em centimorgans. Gaffney (2001) sugere que os QTLs sejam uniformemente distribuídos por todo o genoma, de comprimento L . Dessa forma, *a priori* para a localização do j -ésimo QTL é $\pi(\lambda_j) = 1/L$. A distribuição da posição dos p QTLs, Λ^p , é independente e identicamente distribuída (i.i.d.), dada por

$$\Lambda_j = \pi(\lambda_1, \lambda_2, \dots, \lambda_j) = L^{-p}.$$

A priori uniforme indica que, inicialmente, há pouca informação sobre a posição do QTL no genoma, pois com relação a essa *priori*, qualquer intervalo de mesmo comprimento tem a mesma área (probabilidade), isto é, a chance de um QTL estar localizado em determinado cromossomo é a mesma em todos os cromossomos de mesmo comprimento.

Média e variância

Admite-se que a média, μ , é normalmente distribuída com média ν e variância τ^2 . Uma alternativa é assumir $\nu = 0$ (SATAGOPAN et al., 1996) ou ainda, $\nu = \bar{y} = \frac{1}{n} \sum_i^n y_i$. Outra possível alternativa seria especificar um valor para a média obtido de experimentos anteriores

similares. Para a variância média, assume-se $\tau^2 = s_y^2$, sendo $s_y^2 = \frac{1}{n-1} \sum_i^n (y_i - \bar{y})^2$. Nesse trabalho, a distribuição *a priori* para a média geral, $\pi(\mu)$ é Normal com média obtida pelos valores fenotípicos, \bar{y} , e variância amostral, s_y^2 , ou seja,

$$\mu \sim N(\bar{y}, s_y^2).$$

A *priori* $N(\nu, \tau^2)$ expressa o fato de que τ é uma confiança ou certeza de que ν é um valor razoável para μ . Quanto maior τ^2 (ou τ), maior é a incerteza feita pela natureza com relação a μ (BOLFARINE; SANDOVAL, 2001).

Para a variância fenotípica, geralmente assume-se uma *priori* gama invertida, IG ; com os parâmetros $\alpha = 3$ e $\beta = s_y^2$, isto é

$$\sigma_y^2 \sim IG(3, s_y^2).$$

A função densidade de probabilidade (fdp) gama com os parâmetros $\alpha \geq 1$ e $\beta > 0$ é uma extensão do modelo exponencial (BUSSAB; MORETIN, 2003). Se $\alpha = 1$, tem-se uma fdp exponencial (β).

Efeitos de aditividade e dominância

Para os efeitos genéticos dos QTLs de aditividade e dominância, geralmente assume-se distribuição *a priori* Beta com parâmetros α e β . Os parâmetros α e β da *priori* Beta devem ser escolhidos de modo que $\pi(\theta)$ expresse o conhecimento *a priori* que se tem sobre θ . Satagopan et al. (1996) e Yandell (2004) sugerem uma *priori* $Beta(\alpha = 2, \beta = 10)$, ou seja

$$a \sim Beta(2, 10),$$

e

$$d \sim Beta(2, 10).$$

No caso particular em que $\alpha = \beta = 1$, tem-se $\pi(\theta) = 1$, $0 < \theta < 1$, ou seja, nesse caso a distribuição uniforme $U(0, 1)$ é escolhida como *priori* para os efeitos de aditividade (a) e dominância (d) (BOLFARINE; SANDOVAL, 2001).

2.2 Material e Métodos

2.2.1 Material genético

Os dados utilizados neste trabalho são provenientes de 256 progênies $F_{2:3}$ de milho, obtidas a partir do cruzamento entre as linhagens L02-03D e L20-01F contrastantes para vários caracteres agrônômicos. Ambas as linhagens foram desenvolvidas no Departamento de Genética da ESALQ/USP, sob a coordenação do Prof. Dr. Cláudio Lopes Souza Júnior. Os dados genotípicos de marcadores moleculares para todas as progênies utilizadas no experimento foram obtidos no Laboratório de Biologia Molecular e Análises Genéticas, situado no Centro de Biologia Molecular e Engenharia Genética (CBMEG), na Universidade Estadual de Campinas/UNICAMP, sob a responsabilidade da Profa. Dra. Anete Pereira de Souza.

Efetuada o cruzamento entre as linhagens contrastantes, foi obtida a geração F_1 . A seguir, foram selecionadas três espigas $F_{1's}$ com base na presença de alelos parentais, sendo, em seguida, três plantas dessa geração autofecundadas, dando origem à geração F_2 . Posteriormente, as plantas da população F_2 foram autofecundadas, e as sementes de suas espigas foram semeadas em linhas isoladas, gerando as progênies $F_{2:3}$, que foram avaliadas em experimentos em látices simples 16×16 com duas repetições em nove ambientes, sendo cada ambiente correspondente à combinação local \times ano agrícola. Cada parcela foi constituída por uma linha de 4m de comprimento, com 0,80m de distância entre linhas e 0,20m entre plantas dentro da linha. Foram obtidas 256 progênies $F_{2:3}$ para análise. O caráter estudado neste trabalho foi produção de grãos de milho, em gramas planta⁻¹, ajustado para o estande médio por covariância e corrigido para 15% de umidade (BENTO, 2006). O objetivo desse procedimento é diminuir a variabilidade do número médio de plantas por parcela e padronizar a porcentagem de umidade nos grãos.

2.2.2 Mapeamento de QTLs

Com as informações dos dados fenotípicos associadas ao mapa de ligação elaborado, foram identificadas as possíveis ligações entre os marcadores e os QTLs relacionados ao caráter avaliado, sendo para isso utilizado o módulo *Bmapqtl* do aplicativo *QTLCartographer* versão 2.5, para Windows (WANG; BASTEN; ZENG, 2005). O número de QTLs foi determinado pelo Fator de Bayes (ANEXO A). Os gráficos das seqüências RJMCMC, densidades

a posteriori e obtenção dos intervalos de credibilidade foram feitos no *software* livre *R* versão 2.0.1, para Windows (ANEXOS B e C).

Para o mapeamento de QTLs foram utilizadas as médias da produção de grãos. As associações entre marcadores e QTLs foram realizadas através dos métodos de mapeamento por intervalo composto (CIM) e Bayesiano (BIM). Neste trabalho os dois procedimentos foram utilizados por se tratarem de métodos baseados em análise univariada, permitindo, assim, uma comparação entre eles.

2.2.3 Método 1: Mapeamento por Intervalo Composto

A seleção dos marcadores que possuem maior associação com o caráter produção de grãos (cofatores) foi efetuada através do procedimento *stepwise*, de acordo com o modelo de regressão linear múltipla. Após a seleção dos cofatores que seriam incorporados no modelo, para testar a presença de QTLs foi utilizado o teste da razão de verossimilhanças, correspondente à estatística do teste LR (eq. 43):

$$LR = -2\ln \left[\frac{\text{máx. } L(H_0)}{\text{máx. } L(H_a)} \right] \quad (43)$$

em que $L(H_0)$ é o valor da função de verossimilhança sob a hipótese de nulidade para os efeitos de aditividade e dominância para o genótipo do QTL, ou seja, $H_0 : a = 0, d = 0$; $L(H_1)$ é o valor da função de verossimilhança sob a hipótese alternativa, para os efeitos de aditividade e dominância para o genótipo do QTL, $H_1 : a \neq 0, d \neq 0$.

Assim, testando-se

H_0/H_1 : o valor do teste significativo indica presença de efeito aditivo e/ou de dominância. O efeito do QTL será considerado significativo, onde o valor do LR for o máximo dos que ultrapassarem o limite crítico correspondente. Adotou-se o valor crítico de LR igual a 11,5, comum em trabalhos de mapeamento de QTLs na cultura de milho (ou analogamente, $LODscore$ igual a 2,5).

Foi declarada a presença do QTL em uma dada posição do cromossomo quando o LR calculado ao longo do cromossomo atingiu o valor máximo dentre os que ultrapassaram $LR = 11,5$. Nos casos em que houve mais de um pico com valor de LR superior ao limite crítico em um mesmo intervalo, o QTL foi declarado presente na posição em ocorreu maior valor. Acima do limite, considera-se que cada pico contém as estimativas dos efeitos de aditividade (a) e de dominância (d) do QTL bem como sua posição no genoma (*locus*) (ZENG, 1994).

Para identificar a linhagem doadora do alelo favorável à expressão do caráter, observou-se o sinal do efeito aditivo estimado para cada QTL. Sendo as linhagens contrastantes para vários caracteres, o efeito aditivo estimado com sinal positivo indicou que a origem do alelo contribuindo para maior produção de grãos foi a linhagem parental mais vigorosa, L02-03D. Por outro lado, o efeito aditivo com sinal negativo indicou que a origem do alelo favorável para maior produção de grãos foi a linhagem parental menos vigorosa, L20-01F.

2.2.3.1 Estimativas dos parâmetros genéticos

Para estimar as variâncias genética e fenotípica, foi realizada a análise de variância, seguindo o delineamento experimental em látices simples 16×16 , com efeitos aleatórios, exceto a média.

Para a realização da análise de variância no delineamento em látices, deve-se primeiramente fazer todas as análises individuais, isto é, a análise para cada local. A seguir deve-se examinar as grandezas dos quadrados médios dos resíduos (QMRes). Se forem homogêneos (superior um ao outro em até quatro vezes), todos os locais poderão ser incluídos na análise conjunta sem restrições (BOX, 1954, apud BARBIN, 2003). Em casos contrários, sub-grupos com QMRes homogêneos devem ser organizados e análises conjuntas são feitas para cada sub-grupo ou, alternativamente, podem-se reunir todos esses locais em uma só análise conjunta, porém ajustando-se o número de graus de liberdade.

A Tabela 6 apresenta o esquema da análise de variância conjunta para o experimento em látices no delineamento em blocos casualizados em a ambientes, p progênies e r repetições, considerando-se homogêneos os QMRes obtidos em cada análise individual. Maiores detalhes são apresentados por Bento (2006). É interessante notar que, um delineamento em látice com duas repetições ($r = 2$) constitui o reticulado duplo (*double lattice*), também chamado látice simples (*simple lattice*).

Tabela 6 - Esquema da análise de variância para o experimento em látices no delineamento em blocos casualizados em a ambientes, p progênies, r repetições

Fontes de variação	G.L.	Q.M.	E[Q.M.]
Ambientes (A)	$a - 1$		
Progênies (P)	$p - 1$	Q_1	$\sigma^2 + r\sigma_{GA}^2 + ra\sigma_G^2$
P \times A	$(a - 1)(p - 1)$	Q_2	$\sigma^2 + r\sigma_{GA}^2$
Resíduo Médio	$a(p - 1)(r - 1)$	Q_3	σ^2

Fonte: Barbin (2003)

A partir das esperanças e dos quadrados médios da análise de variância, foram estimados os seguintes parâmetros genéticos:

Variância genética, $\hat{\sigma}_G^2$:

$$\hat{\sigma}_G^2 = \frac{Q_1 - Q_2}{ra}, \quad (44)$$

Variância fenotípica entre as médias de progênies, $\hat{\sigma}_F^2$:

$$\hat{\sigma}_F^2 = \frac{Q_1}{ra} = \hat{\sigma}_G^2 + \frac{\hat{\sigma}_{GA}^2}{a} + \frac{\hat{\sigma}^2}{ra}, \quad (45)$$

Coefficiente de herdabilidade do caráter, \hat{h}^2 (BARBIN, 1993):

$$\hat{h}^2 = \frac{\hat{\sigma}_G^2}{\hat{\sigma}_F^2}. \quad (46)$$

Variância genética explicada pelo i -ésimo QTL em uma população F_2 , $\hat{\sigma}_G^2 Q_i$:

$$\hat{\sigma}_G^2 Q_i = \frac{\hat{a}_i^2}{2} + \frac{\hat{d}_i^2}{4}, \quad (47)$$

sendo \hat{a}_i^2 e \hat{d}_i^2 os efeitos aditivo e de dominância estimados para o QTL Q_i .

Proporção da variância fenotípica explicada pelo i -ésimo QTL, $R_F^2 Q_i$:

$$R_F^2 Q_i = \left(\frac{\hat{\sigma}_G^2 Q_i}{\hat{\sigma}_F^2} \right), \quad (48)$$

sendo $\hat{\sigma}_G^2 Q_i$ é a variância genotípica estimada para o i -ésimo QTL.

Proporção da variância genotípica explicada pelo i -ésimo QTL, $R_G^2 Q_i$:

$$R_G^2 Q_i = \left(\frac{\hat{R}_F^2 Q_i}{\hat{h}^2} \right), \quad (49)$$

em que \hat{h}^2 é o coeficiente de herdabilidade do caráter.

Proporção da variância fenotípica total, $R_F^2 Total$ (eq. 50):

$$R_F^2 Total = \sum_{i=1}^n R_F^2 Q_i. \quad (50)$$

Proporção da variância genotípica total, $R_G^2 Total$ (eq. 51):

$$R_G^2 Total = \sum_{i=1}^n R_G^2 Q_i. \quad (51)$$

Grau de dominância para cada QTL mapeado, $\hat{G}D_i$:

$$\hat{G}D_i = |\hat{d}_i|/|\hat{a}_i|. \quad (52)$$

sendo \hat{a}_i e \hat{d}_i são as estimativas, em módulo, dos efeitos de aditividade e dominância do i -ésimo QTL.

Estimados os graus de dominância para cada QTL, pode ser obtida a estimativa do grau médio de dominância, $\hat{G}MD$, obtido pela média dos $\hat{G}D_i$, ponderada pelo coeficiente de determinação, R_G^2 , dos efeitos individuais de cada QTL. Para $\hat{G}D < 0,2$, a ação do QTL foi considerada aditiva; para $0,2 \leq \hat{G}D < 0,8$, como de dominância parcial; $0,8 \leq \hat{G}D < 1,2$, de dominância completa; e $\hat{G}D \geq 1,2$, de sobredominância (GARCIA, 2003; MANGOLIN, 2002).

2.2.4 Método 2: Mapeamento por Intervalo Bayesiano

Neste trabalho, a abordagem Bayesiana de mapeamento de QTLs considerou o número de QTLs (P), e, portanto, o comprimento do vetor dos efeitos de aditividade, dominância e posição dos QTLs aleatórios. Distribuições *a priori* foram especificadas para os efeitos de aditividade, dominância, λ_j e do erro. Para os efeitos genéticos de aditividade e dominância foram assumidas *priori* Beta(2,10) e, para o número de QTLs, Poisson(3). Para a

média geral admitiu-se *priori* $N(\bar{y}, s_y^2)$ e para a variância Gama Invertida(3, s_y^2) (SATAGOPAN et al., 1996; GAFFNEY, 2001). Além disso, realizando-se o mapeamento de QTLs em produção de grãos de milho, o número de QTLs foi determinado pelo Fator de Bayes (SORENSEN; GIANOLA, 2002).

Visto que a distribuição conjunta *a posteriori* para os parâmetros do modelo possui forma analítica complexa, resumos tais como a média e intervalos de credibilidade para os parâmetros de interesse apresentam cálculos inviáveis. Desse modo, foram empregados métodos computacionalmente intensivos (algoritmo Monte Carlo via Cadeias de Markov, MCMC), para a obtenção de uma amostra da distribuição conjunta *a posteriori* e a partir dela fazer inferências sobre os parâmetros de interesse, ou seja, resumos para os efeitos aditivo, dominância e posições dos QTLs.

A suposição de o número de QTLs ser desconhecido, tornou a obtenção da amostra da distribuição conjunta *a posteriori* ligeiramente mais complicada. Uma solução para esse problema esteve na utilização dos algoritmos MCMC com saltos reversíveis, “*Reversible Jump MCMC*” (GREEN, 1995).

Para a obtenção de uma amostra da distribuição conjunta *a posteriori* $\pi(\boldsymbol{\theta}|\mathbf{y})$ foi utilizado o módulo *Bmapqtl* do programa *QTLCartographer*, versão 2.5 para Windows. Foi gerada uma cadeia de Markov com 400.000 iterações, “*burn-in*” de 2.000, “*pré burn-in*” de 1.000 e “*thin*” de 100 iterações, obtendo-se uma amostra resultante de tamanho 15.315.

Após detectados os QTLs foram calculadas as estimativas de seus efeitos de aditividade e dominância, bem como o tipo de interação alélica a eles associados (grau de dominância).

Após se realizar o mapeamento de QTLs pelos dois métodos, (mapeamento por intervalo Composto e mapeamento por intervalo Bayesiano), os resultados foram finalmente comparados.

2.3 Resultados e Discussão

2.3.1 Resultados obtidos pelo método 1 (CIM)

A seleção dos cofatores foi efetuada pelo procedimento *stepwise* implementado no aplicativo *QTLCartographer*, identificando dezesseis marcadores que foram incorporados como covariáveis. As associações entre marcadores e QTLs foram consideradas significativas quando

o LR calculado ao longo do cromossomo atingiu o valor máximo dentre os que ultrapassaram o limiar crítico $LR = 11,5$. Foram detectados dez QTLs para produção de grãos, distribuídos nos cromossomos 3 (três QTLs), 5 (cinco QTLs), e 7 (dois QTLs). A localização, efeitos aditivos e de dominância, assim como os R_F^2 e R_G^2 dos QTLs estão apresentadas na Tabela 7.

As progênies $F_{2:3}$ avaliadas apresentaram produção média de 43,91 gramas planta⁻¹. As variâncias genética ($\hat{\sigma}_G^2$) e fenotípica ($\hat{\sigma}_F^2$) estimadas para o caráter produção de grãos foram 61,88 e 71,29 (g planta⁻¹)², respectivamente. A estimativa do coeficiente de herdabilidade (\hat{h}^2) foi de 86,8%, indicando alta herdabilidade para o caráter (Tabela 8).

Os efeitos aditivos (a) dos QTLs para produção de grãos variaram de -0,47 a 2,48, enquanto os de dominância (d) variaram de -0,57 a 2,36 (Tabela 7).

A maior parte dos alelos que contribui de forma favorável para o caráter é proveniente da linhagem parental mais vigorosa para a produção de grãos, L02-03D. A contribuição através de alelos favoráveis para essa linhagem foi efetiva para oito QTLs. Para os demais QTLs os alelos favoráveis para o aumento da produção de grãos são provenientes da linhagem parental menos vigorosa, L20-01F. Isso pode ser observado pelo sinal do efeito aditivo estimado para cada QTL.

A proporção da variância fenotípica (R_F^2) explicada pelos QTLs variou de 0,94%, para qprod7a a 2,47% para qprod5e. Em contrapartida, a proporção da variância genética (R_G^2) variou de 1,09% a 2,87%, explicada, respectivamente, pelos mesmos QTLs qprod7a e qprod5e. Em conjunto, os dez QTLs detectados para produtividade de grãos, explicaram 17,09% da variância fenotípica e 19,86% da variância genética do caráter.

As estimativas dos efeitos genéticos e dos valores de $|d|/|a|$ para cada QTL indicaram a ocorrência de diferentes tipos de interações alélicas, mostrando que a interação gênica predominante foi de dominância parcial (4 QTLs), seguida por dominância completa (3 QTLs), sobredominância (2 QTLs) e aditiva (1 QTL). O grau médio de dominância estimado para o caráter produção de grãos considerando todos os QTLs mapeados ponderados por seus respectivos R_G^2 (obtidos através da eq. 51) foi de 1,12, evidenciando que esse caráter apresenta em média dominância completa, sendo esse o tipo de efeito mais importante para a produtividade de grãos.

A Tabela 7 apresenta as estimativas para a localização, valores do LR para o efeito do QTL, efeitos genéticos, tipo de interação alélica e proporção da variância fenotípica e genética explicadas pelos QTLs associados a produção de grãos, detectados pelo mapeamento

por intervalo composto. Os nomes dos QTLs são compostos por código indicando o caráter correspondente (qprod - QTL associado a produção de grãos), seguido por número indicando o cromossomo no qual ele se encontra e por letra identificando os vários QTLs localizados em um mesmo cromossomo.

Tabela 7 - Mapeamento de QTLs por intervalo composto para o caráter produção de grãos. Localização, valores do LR para o efeito do QTL, efeitos genéticos, tipo de interação alélica e proporção da variância fenotípica (R_F^2) e genética (R_G^2) individual e total explicadas pelos QTLs

QTL	Crom. ^a	Posição		LR	Efeitos			$G\hat{D}^b$	Direção	$\hat{R}_F^2\%$ ^c	$\hat{R}_G^2\%$ ^d	
		cM	Intervalo		\hat{a}	\hat{d}	$ \hat{d} / \hat{a} $					
qprod3a	3	24,72	B1904-P029	13,56	1,39	1,14	0,82	DC	L02-03D	1,81	2,10	
qprod3b	3	31,21	P029-U1025	12,47	2,48	2,36	0,95	DC	L02-03D	1,57	1,83	
qprod3c	3	49,02	U1025-P053	13,81	1,68	1,00	0,59	DP	L02-03D	2,34	2,72	
qprod5a	5	48,63	D10-B1208	14,28	1,22	0,73	0,60	DP	L02-03D	1,23	1,43	
qprod5b	5	55,71	B1208-M0282	15,54	1,20	1,05	0,87	DC	L02-03D	1,39	1,61	
qprod5c	5	70,74	M0282-U1524	17,30	1,65	-0,01	0,01	A	L02-03D	1,91	2,22	
qprod5d	5	97,62	U1524-U2013	14,33	1,53	-0,57	0,37	DP	L02-03D	1,75	2,03	
qprod5e	5	123,19	B2305-U1792	19,99	1,85	-0,44	0,24	DP	L02-03D	2,47	2,87	
qprod7a	7	77,68	B657-U1112	13,60	-0,47	1,49	3,16	SD	L20-01F	0,94	1,09	
qprod7b	7	126,72	B1805-P116	14,08	-0,41	2,11	5,18	SD	L20-01F	1,68	1,95	
							GMD ^e	1,12	DC	R^2 Total	17,09	19,86

^aCromossomo em que foi detectado o QTL; ^bGD (grau de dominância): A=aditiva ($|\hat{d}|/|\hat{a}| < 0,2$), DP=dominância parcial ($0,2 \leq |\hat{d}|/|\hat{a}| < 0,8$), DC=dominância completa ($0,8 \leq |\hat{d}|/|\hat{a}| < 1,2$), SD=sobredominância ($|\hat{d}|/|\hat{a}| \geq 1,2$); ^c $R_F^2\%$: porcentagem da variação fenotípica explicada pelo QTL; ^d $R_G^2\%$: porcentagem da variação genética explicada pelo QTL; ^e $G\hat{M}D$ (grau médio de dominância) = $\frac{\sum_i G\hat{D}_i \hat{R}_{G_i}^2}{\sum_i \hat{R}_{G_i}^2}$.

As estimativas para a média, variância genética, variância fenotípica e do coeficiente de herdabilidade para produção de grãos da população $F_{2:3}$ são apresentadas na Tabela 8:

Tabela 8 - Valor médio e estimativas das variâncias genética, fenotípica e do coeficiente de herdabilidade para produção de grãos da população $F_{2:3}$

Parâmetros	Produção de grãos (g planta⁻¹)
Média	43,91
$\hat{\sigma}_G^2$	61,88
$\hat{\sigma}_F^2$	71,29
h^2	0,868

Os QTLs detectados pelo mapeamento por intervalo composto, localizados no genoma do milho podem ser vistos através da Figura 11. A Figura 12 mostra os valores críticos do teste LR para a declaração de QTLs em cada cromossomo.

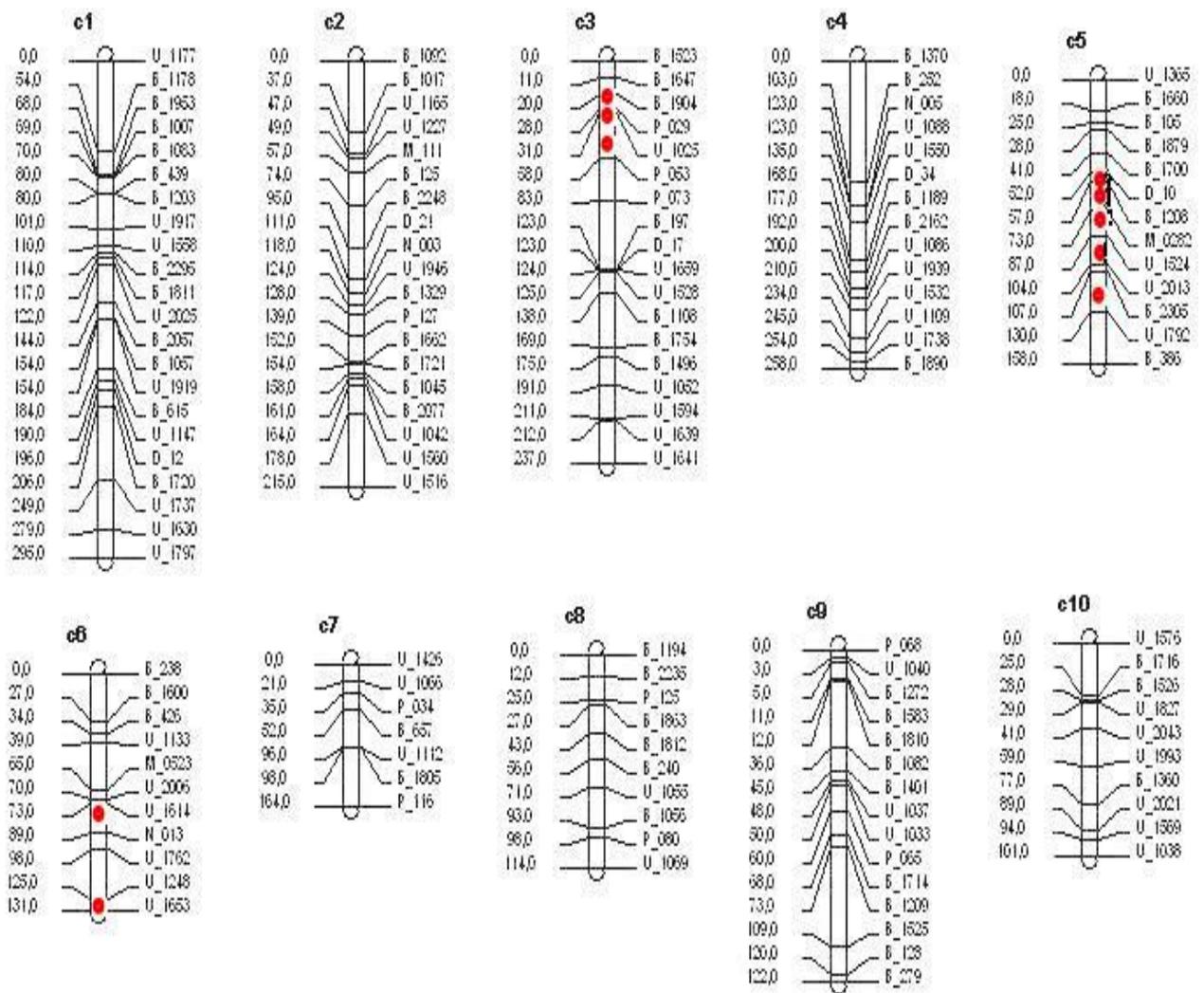


Figura 11 - Mapa genético dos dez cromossomos de milho com 139 marcas e com os dez QTLs obtidos pelo mapeamento por intervalo composto. As posições dos QTLs nos cromossomos estão indicadas pelos círculos vermelhos

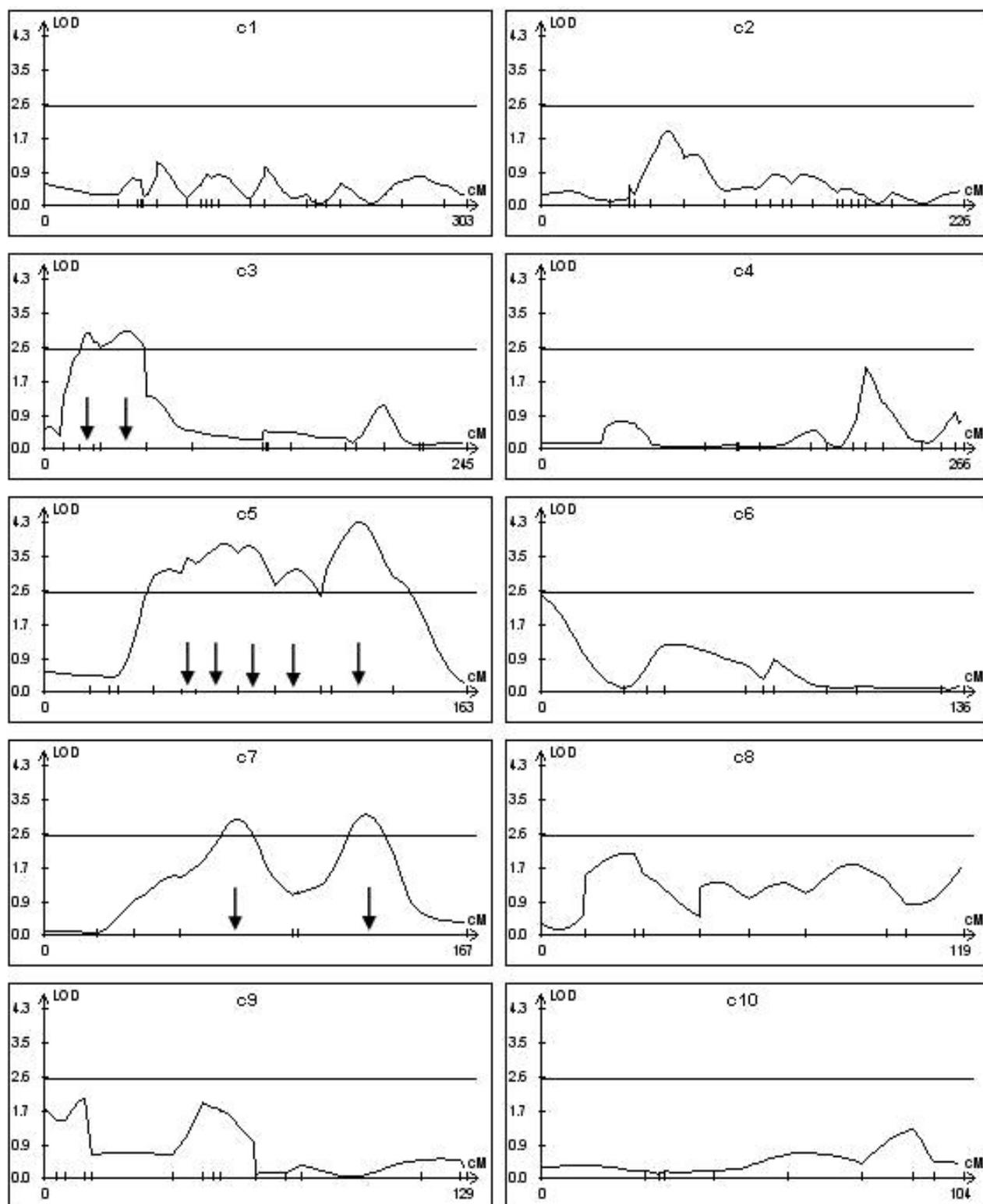


Figura 12 - Mapeamento de QTLs associados a produção de grãos, obtido pelo mapeamento por intervalo composto. As posições dos QTLs nos cromossomos estão indicadas pelas setas

2.3.2 Resultados obtidos pelo método 2 (BIM)

A distribuição conjunta *a posteriori* foi obtida através do processo de amostragem “*Reversible Jump MCMC*”, em que foram encontrados modelos contendo de zero a treze QTLs (Figura 13). O gráfico em barras mostra que as maiores frequências estão entre os modelos com três, quatro e cinco QTLs.

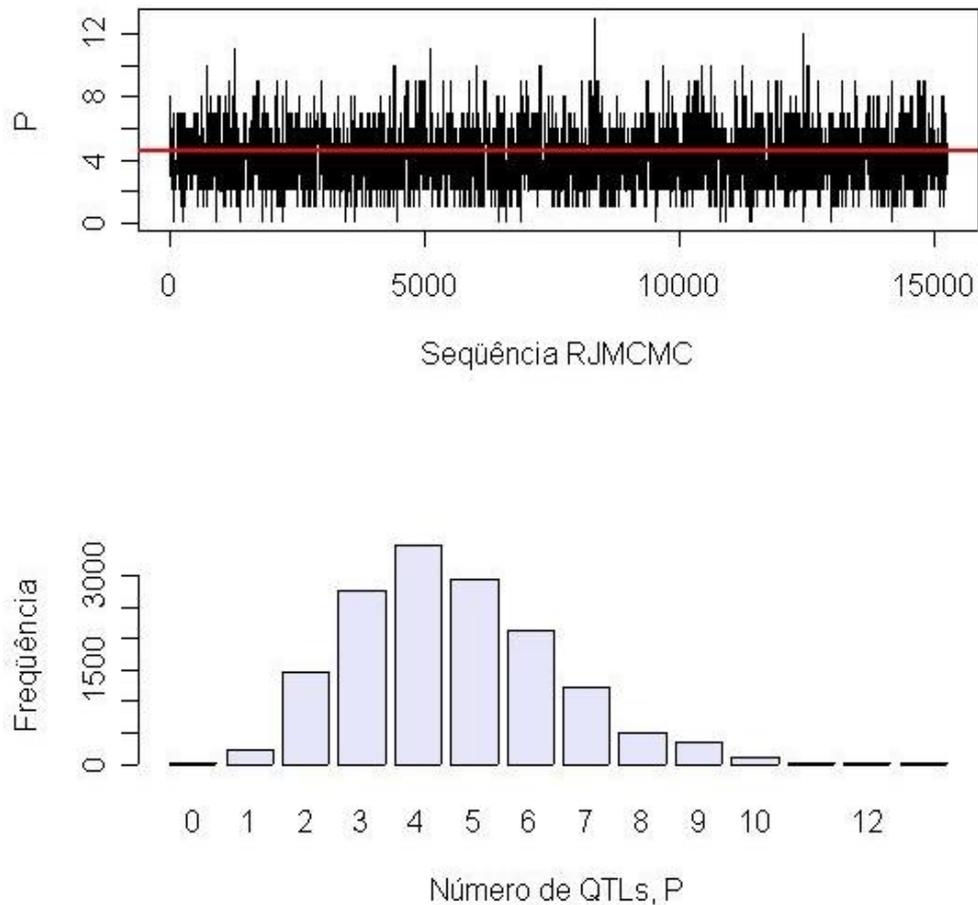


Figura 13 - Seqüências RJMCMC e densidades *a posteriori* marginais, $P(\theta|\mathbf{y})$, para o número de QTLs

Através da Figura 13, a seqüência “*Reversible Jump MCMC*” e as frequências para o número de QTLs (P) podem ser verificadas. O número médio de QTLs igual a 4,6 é representado pela linha vermelha horizontal. A Figura 14 mostra as distribuições *a priori* e *a posteriori* para P .

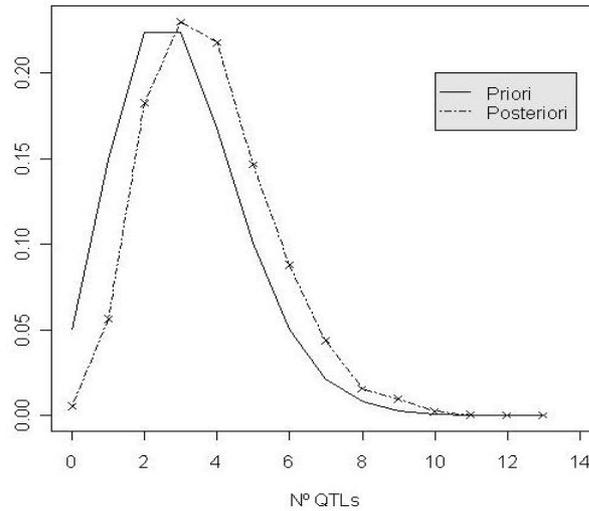


Figura 14 - Densidades *a priori* e *a posteriori* para o número de QTLs

Para declarar o número de QTLs no modelo, foi adotado o critério do Fator de Bayes (YANDELL et al., 2004). Calculados os Fatores de Bayes (ANEXO B), o valor $B(M_5, M_4) = 1,264$ indicou evidência a favor do modelo M_5 em relação a M_4 , levando-se a conclusão de que o modelo com cinco QTLs se ajustou melhor aos dados.

Considerando-se o modelo com cinco QTLs, M_5 , foram calculados para cada QTL sua posição no cromossomo (λ), seus efeitos genéticos: efeito aditivo (a) e de dominância (d) e o tipo de interação alélica (GD) ocorrida, assim como os R_{F*}^2 e R_{G*}^2 apresentados na Tabela 9. Os nomes dos QTLs são compostos por código indicando o caráter correspondente seguido por número indicando o cromossomo no qual se encontra o QTL, por letra identificando os vários QTLs localizados no mesmo cromossomo e por asterisco (*), indicando que o QTL foi detectado pelo método de mapeamento por intervalo Bayesiano. Dessa forma, observa-se que QTL nomeado por qprod5 significa presença de QTL para o caráter produção de grãos localizado no cromossomo cinco. As letras a, b identificam os vários QTLs localizados no mesmo cromossomo cinco. A Tabela 10 apresenta os intervalos de credibilidade de 95% para a localização (λ) e os efeitos aditivo (a) e de dominância (d).

As progênies $F_{2:3}$ avaliadas apresentaram produção média de 43,84 gramas planta⁻¹. O coeficiente de herdabilidade (\hat{h}^2) foi de 15%, indicando baixa herdabilidade para o caráter (Tabela 11).

Os cinco QTLs associados a produção de grãos, mapeados através da abordagem Bayesiana, estão distribuídos nos cromossomos 1 (um QTL), 3 (um QTL), 5 (dois QTLs), e 9 (um QTL). Os efeitos aditivos (a) dos QTLs para produção de grãos variaram de -0,23 a 0,92, enquanto os de dominância (d) variaram de -0,16 a 0,34, para os QTLs qprod5b* e qprod9a*

respectivamente.

Os valores de $|d|/|a|$ para cada QTL indicaram a ocorrência de diferentes tipos de interações alélicas: dominância parcial (dois QTLs), aditiva (dois QTLs), e sobredominância (um QTL).

A proporção da variância fenotípica (R_{F*}^2) explicada pelos QTLs variou de 0,09% a 0,69%, enquanto a proporção da variância genética (R_{G*}^2) variou de 0,62% a 4,69%, explicada, respectivamente, pelos QTLs qprod9a* e qprod5b*. Em conjunto, os cinco QTLs detectados para produção de grãos, explicaram 1,93% da variância fenotípica e 13,06% da variância genética do caráter.

As estimativas dos efeitos genéticos e dos valores de $|d|/|a|$ para cada QTL indicaram a ocorrência de diferentes tipos de interações alélicas, mostrando que as interações gênicas predominantes foram de dominância parcial (2 QTLs) e aditiva (2 QTLs), seguidas por sobredominância (1 QTL). O grau médio de dominância estimado para o caráter produtividade de grãos considerando todos os QTLs mapeados ponderados por seus respectivos (R_{G*}^2) foi de 0,28, evidenciando que o caráter produção de grãos apresenta em média dominância parcial.

A maior parte dos alelos que contribui de forma favorável para o caráter é proveniente da linhagem parental L02-03D, que apresentou maior produção de grãos. A contribuição através de alelos favoráveis para essa linhagem foi efetiva para quatro dos cinco QTLs mapeados.

A Tabela 9 apresenta as estimativas para a localização, efeitos genéticos, tipo de interação alélica e proporção das variâncias fenotípica e genética explicadas pelos QTLs associados a produção de grãos, detectados pelo mapeamento por intervalo Bayesiano.

Tabela 9 - Mapeamento de QTLs por intervalo Bayesiano para o caráter produção de grãos. Localização, efeitos genéticos, tipo de interação alélica e proporção da variância fenotípica (R_{F*}^2) e genética (R_{G*}^2) individual e total explicadas pelos QTLs

QTL	Posição		Intervalo	Efeitos			G^D ^b	Direção	\hat{R}_{F*}^2 % ^c	\hat{R}_{G*}^2 % ^d	
	Crom. ^a	cM		\hat{a}	\hat{d}	$ \hat{d} / \hat{a} $					
qprod1a*	1	106,02	U1917-U1558	0,51	0,19	0,38	DP	L02-03D	0,23	1,53	
qprod3a*	3	107,34	P073-B197	0,81	0,09	0,11	A	L02-03D	0,54	3,67	
qprod5a*	5	75,52	M0282-U1524	0,66	0,25	0,37	DP	L02-03D	0,38	2,56	
qprod5b*	5	98,05	U1524-U2013	0,92	-0,16	0,18	A	L02-03D	0,69	4,69	
qprod9a*	9	64,84	P065-B1714	-0,23	0,34	1,45	SD	L20-01F	0,09	0,62	
						GMD ^e	0,28	DP	R^2 Total	1,93	13,06

^aCromossomo em que foi detectado o QTL; ^bGD(grau de dominância); ^c R_{F*}^2 %: porcentagem da variação fenotípica explicada pelo QTL; ^d R_{G*}^2 %: porcentagem da variação genética explicada pelo QTL; ^eGMD (grau médio de dominância), definidos no procedimento anterior.

Os intervalos de credibilidade para os parâmetros genéticos dos QTLs mapeados (ANEXO D), bem como de sua localização no genoma, são apresentados na Tabela 10.

Tabela 10 - Intervalos de credibilidade para os efeitos genéticos, (a) e (d) e localização (λ) dos QTLs controladores do caráter produção de grãos

QTL ^{a*}	Posição	Efeitos	
	λ	a	d
qprod1a*	(12,56; 276,47)	(-1,95; 3,10)	(-1,61; 2,19)
qprod3a*	(11,03; 275,27)	(-2,10; 3,70)	(-1,93; 2,37)
qprod5a*	(4,62; 153,48)	(-2,36; 3,85)	(-1,93; 2,56)
qprod5b*	(3,64; 240,11)	(-2,21; 3,76)	(-2,64; 1,96)
qprod9a*	(2,95; 141,29)	(-2,64; 3,38)	(-2,03; 2,73)

A Tabela 11 apresenta as estimativas da média geral, coeficiente de herdabilidade e variância genotípica, obtidas através do método Bayesiano.

Tabela 11 - Valor médio, estimativas da variância genética e do coeficiente de herdabilidade para produção de grãos da população $F_{2:3}$, obtidos pelo procedimento Bayesiano

Parâmetros	Produção de grãos (g planta ⁻¹)
Média	43,84
$\hat{\sigma}_G^2$	10,54
$\hat{\sigma}_F^2$	61,30
h^2	0,15

Os QTLs detectados pelo mapeamento por intervalo Bayesiano, localizados no genoma do milho podem ser vistos através da Figura 15. As Figuras 16 a 20 mostram seqüências “*Reversible Jump MCMC*” e densidades marginais *a posteriori*, $\pi(\mathbf{a}_j|\mathbf{y})$, $\pi(\mathbf{d}_j|\mathbf{y})$ e $\pi(\boldsymbol{\lambda}_j|\mathbf{y})$ associadas aos efeitos de aditividade (a_j), de dominância (d_j) e de sua localização no cromossomo (λ_j), $j = 1, \dots, 5$, para cada um dos cinco QTLs mapeados. As densidades marginais foram obtidas a partir da *posteriori* conjunta, $\pi(\boldsymbol{\theta}|\mathbf{y})$, gerada pelo procedimento RJMCMC. Os histogramas representam densidades *a posteriori* de cada parâmetro, suavizada pelas linhas azuis.

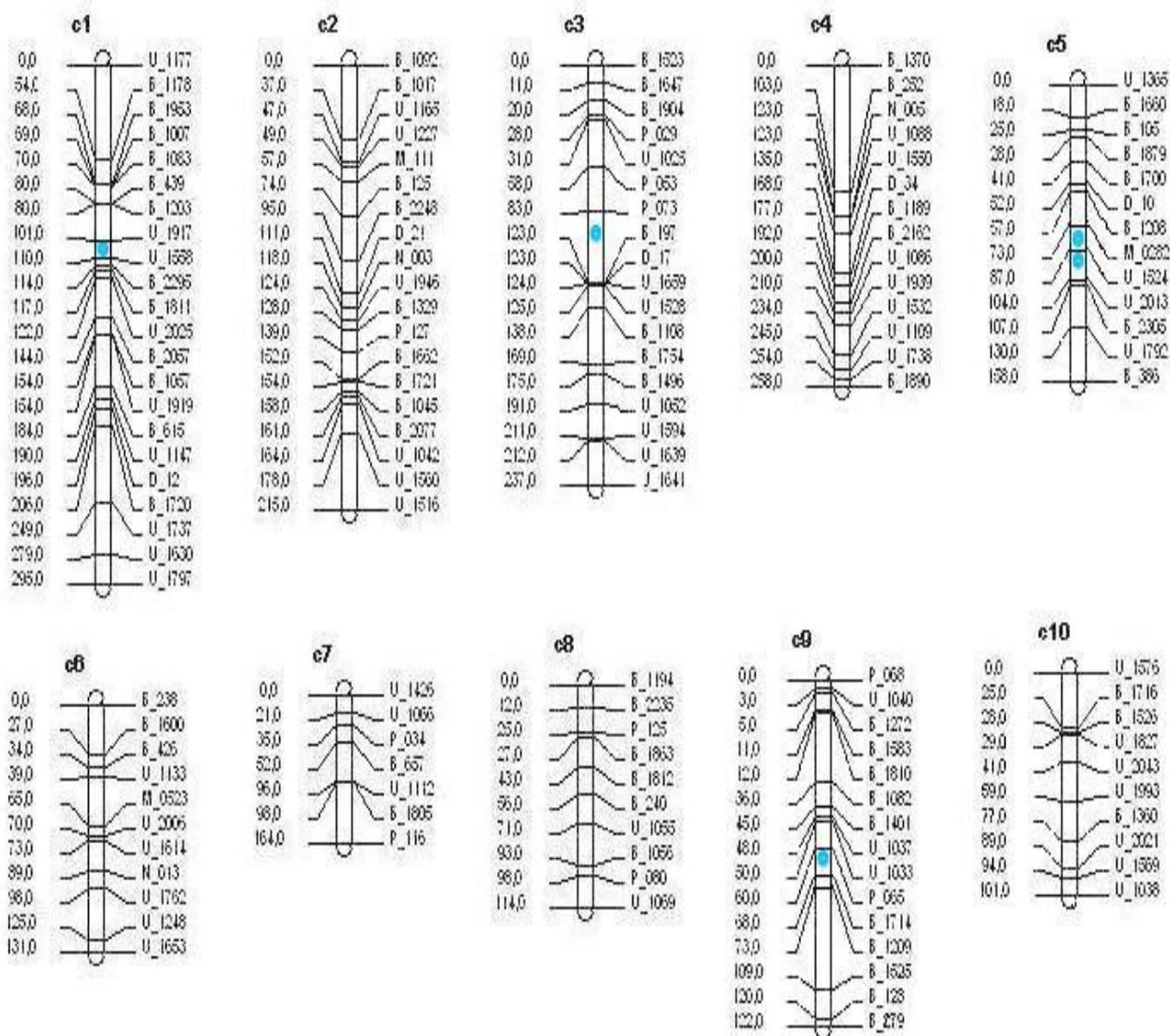


Figura 15 - Mapa genético dos dez cromossomos de milho com 139 marcas e com os cinco QTLs obtidos pelo mapeamento por intervalo Bayesiano, representados pelos círculos azuis

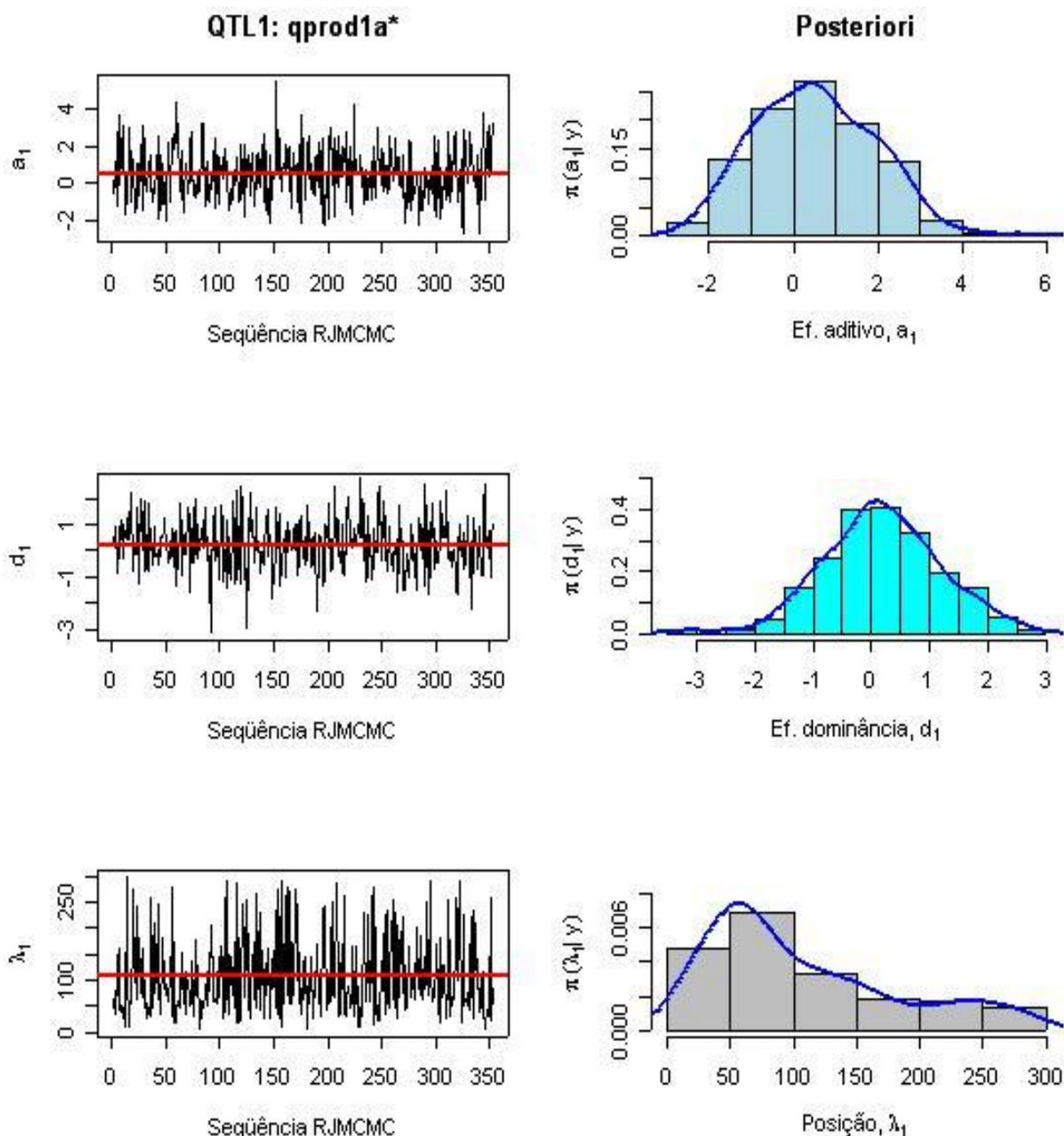


Figura 16 - Seqüências RJMCMC e densidades *a posteriori* marginais ($\pi(\mathbf{a}_1|\mathbf{y})$, $\pi(\mathbf{d}_1|\mathbf{y})$ e $\pi(\lambda_1|\mathbf{y})$) para os efeitos de aditividade (a_1), de dominância (d_1) e posição (λ_1) do primeiro QTL mapeado, qprod1a*

Através da Figura 16 são observadas as seqüências RJMCMC e as densidades marginais *a posteriori* $\pi(\mathbf{a}_1|\mathbf{y})$, $\pi(\mathbf{d}_1|\mathbf{y})$ e $\pi(\lambda_1|\mathbf{y})$ referentes aos efeitos de aditividade (a_1), dominância (d_1) e localização (λ_1), para o primeiro QTL mapeado, qprod1a*. As médias *a posteriori*, são estimativas para $\hat{a}_1 = 0,51$, $\hat{d}_1 = 0,19$ e $\hat{\lambda}_1 = 106,02$ representados pelas linhas vermelhas horizontais em suas respectivas séries.

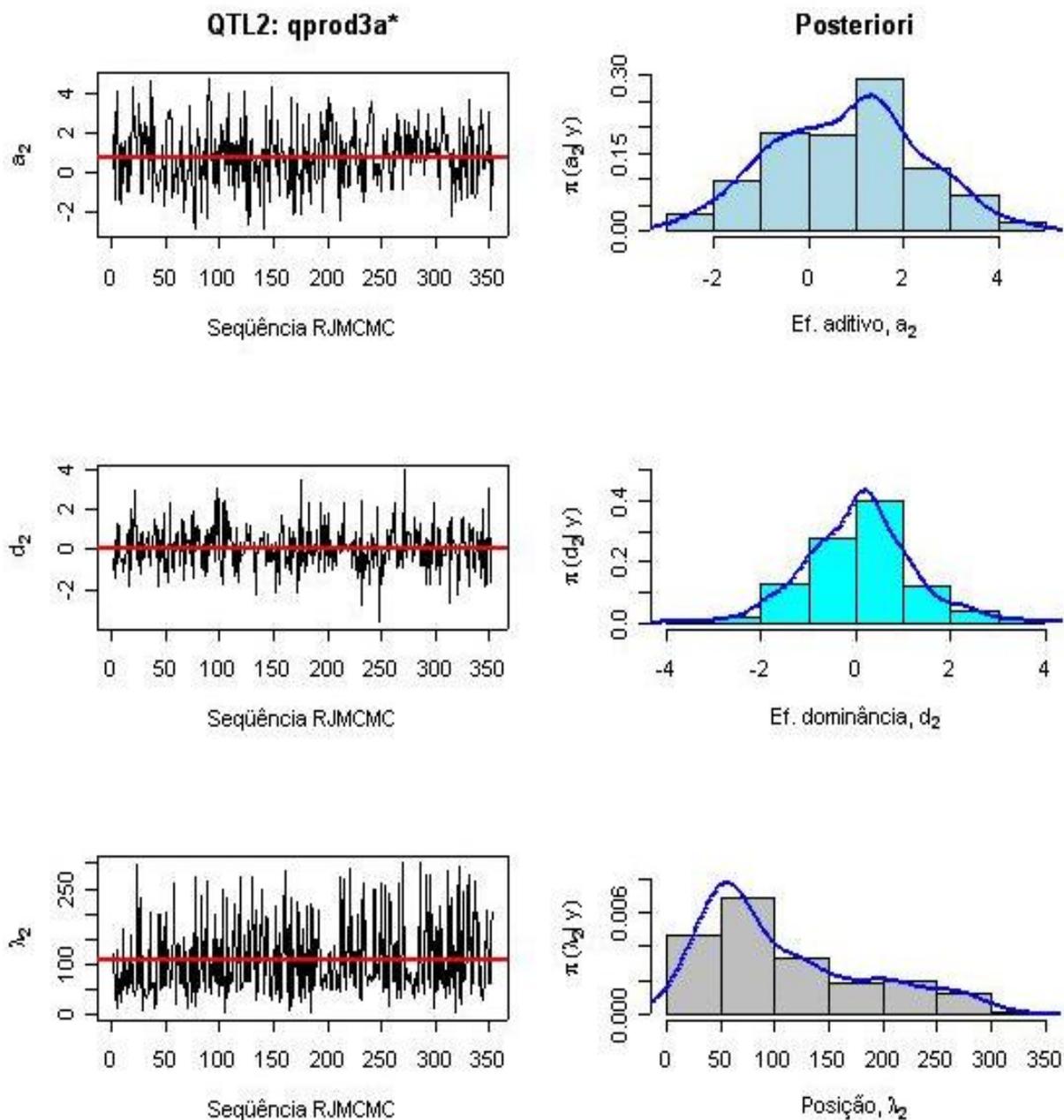


Figura 17 - Seqüências RJMCMC e densidades *a posteriori* marginais ($\pi(a_2|y)$, $\pi(d_2|y)$ e $\pi(\lambda_2|y)$) para os efeitos de aditividade (a_2), de dominância (d_2) e posição (λ_2) do segundo QTL mapeado, qprod3a*

São observados através da Figura 17 as seqüências RJMCMC e os histogramas das densidades marginais *a posteriori* para os efeitos de aditividade (a_2), dominância (d_2) e localização (λ_2), associados ao segundo QTL mapeado, qprod3a*. As estimativas para $\hat{a}_2 = 0,81$, $\hat{d}_2 = 0,09$ e $\hat{\lambda}_2 = 107,34$ estão representados pelas linhas vermelhas horizontais em suas respectivas séries.

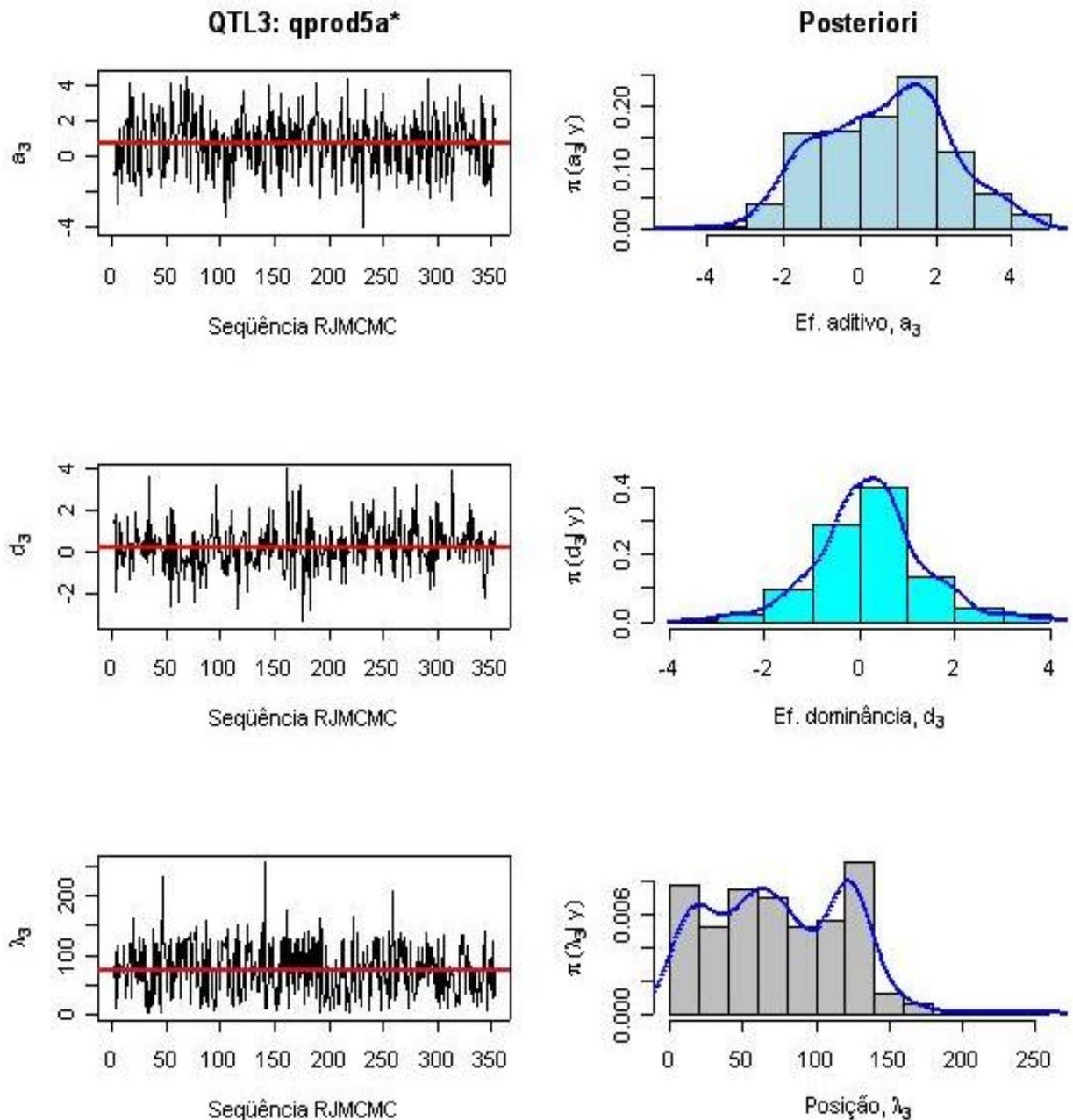


Figura 18 - Seqüências RJMCMC e densidades *a posteriori* marginais ($\pi(a_3|y)$, $\pi(d_3|y)$ e $\pi(\lambda_3|y)$) para os efeitos de aditividade

(a_3), de dominância (d_3) e posição (λ_3) do terceiro QTL mapeado, qprod5a*

A Figura 18 mostra as seqüências RJMCMC e as densidades marginais *a posteriori* referentes aos efeitos de aditividade, dominância e localização, para o terceiro QTL mapeado, qprod5a* e suas estimativas ($\hat{a}_3 = 0,66$, $\hat{d}_3 = 0,25$ e $\hat{\lambda}_3 = 75,52$), representados pelas linhas vermelhas.

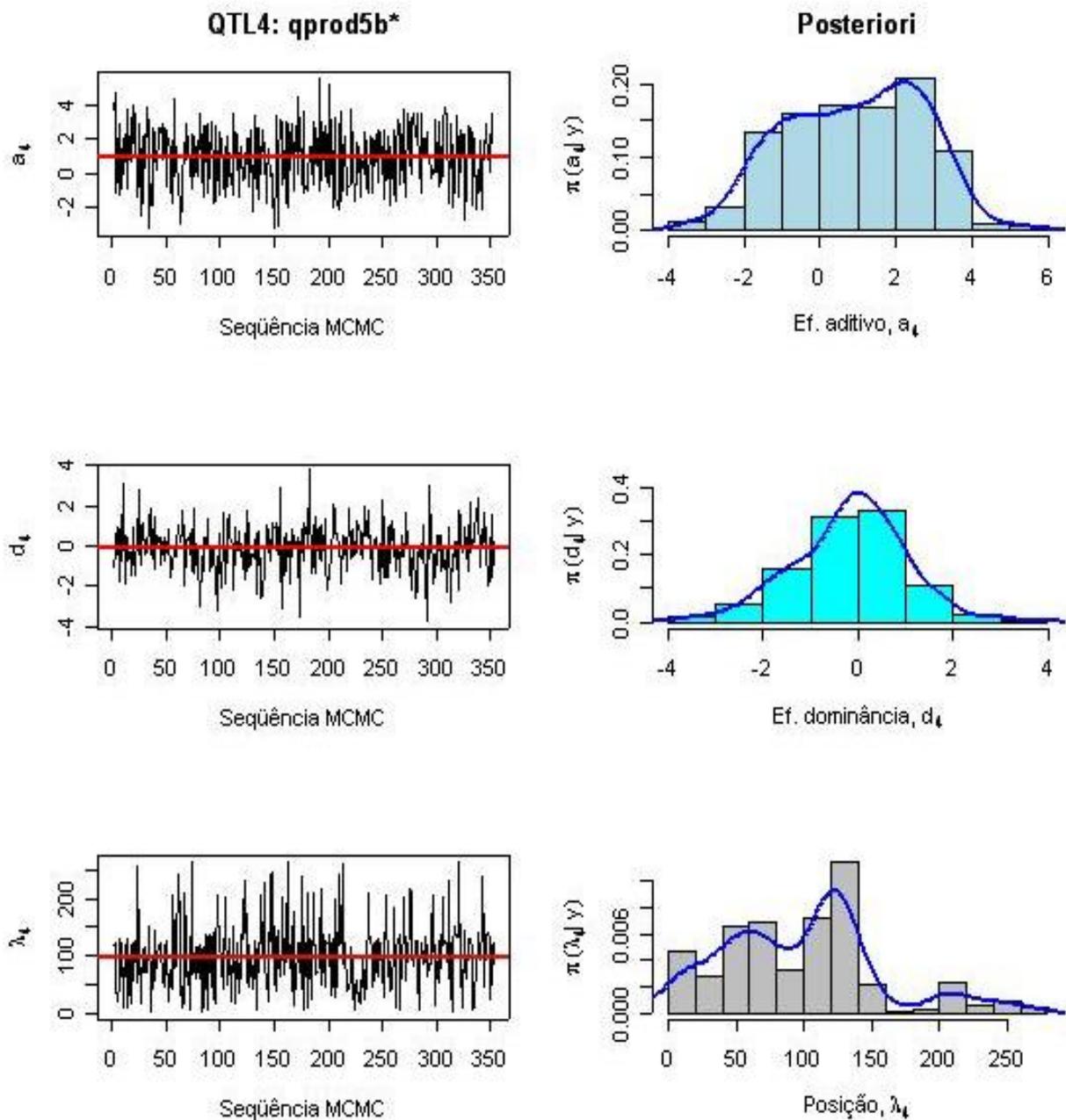


Figura 19 - Seqüências RJMCMC e densidades *a posteriori* marginais ($\pi(\mathbf{a}_4|\mathbf{y})$, $\pi(\mathbf{d}_4|\mathbf{y})$ e $\pi(\lambda_4|\mathbf{y})$) para os efeitos de aditividade (a_4), de dominância (d_4) e posição (λ_4) do quarto QTL mapeado, qprod5a*

Observam-se, pela Figura 19 as seqüências RJMCMC e histogramas das densidades marginais *a posteriori* associadas aos efeitos de aditividade, dominância e localização, para o quarto QTL mapeado, qprod5b*. Suas estimativas são: $\hat{a}_4 = 0,92$, $\hat{d}_4 = -0,16$ e $\hat{\lambda}_4 = 98,05$.

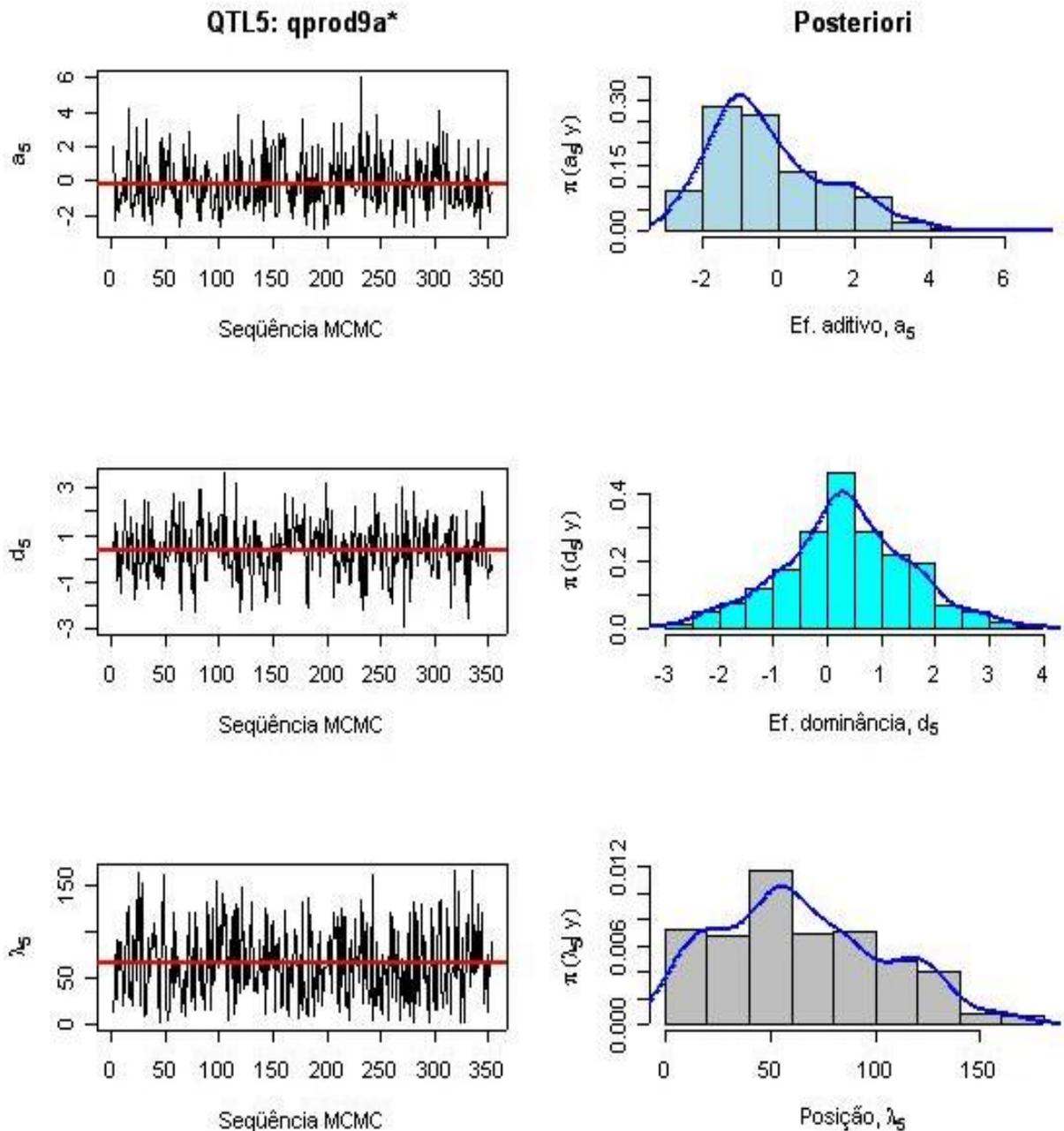


Figura 20 - Seqüências RJMCMC e densidades *a posteriori* marginais ($\pi(\mathbf{a}_5|\mathbf{y})$, $\pi(\mathbf{d}_5|\mathbf{y})$ e $\pi(\lambda_5|\mathbf{y})$) para os efeitos de aditividade (a_5), de dominância (d_5) e posição (λ_5) do quinto QTL mapeado, qprod5c*

Através da Figura 20 são observadas as seqüências as seqüências RJMCMC e histogramas de $\pi(\mathbf{a}_5|\mathbf{y})$, $\pi(\mathbf{d}_5|\mathbf{y})$ e $\pi(\lambda_5|\mathbf{y})$ associadas aos efeitos de aditividade, dominância e localização para o quinto QTL mapeado, qprod9a*. Suas estimativas: $\hat{a}_5 = -0,23$, $\hat{d}_5 = 0,34$ e $\hat{\lambda}_5 = 64,84$.

2.3.3 Comparação dos resultados obtidos pelos métodos CIM e BIM

De acordo com os resultados obtidos nesse trabalho podem-se verificar:

Foram mapeados dez QTLs associados a produção de grãos pelo método CIM e cinco QTLs pelo procedimento BIM, concentrados principalmente entre os cromossomos três e cinco. Destes, apenas dois são coincidentes, (qprod5c = qprod5a*) e (qprod5d = qprod5b*), ambos localizados do cromossomo cinco. Os demais QTLs mapeados podem representar apenas parte do número total de QTLs que controlam o caráter nessa população. Os QTLs mapeados, em sua maioria, apresentaram efeito de dominância;

Os dez QTLs mapeados pelo método CIM explicam conjuntamente $R_G^2 = 19,86\%$ da variação genética e $R_F^2 = 17,09\%$ da variação fenotípica para o caráter produção de grãos. Em contrapartida, os cinco QTLs mapeados pelo procedimento BIM, apenas $R_{G*}^2 = 13,06\%$ e $R_{F*}^2 = 1,93\%$ das variações genética e fenotípica, respectivamente. Em ambos os métodos o grau médio de dominância (GMD) foi de dominância: completa pelo CIM e parcial pelo BIM. Pelas estimativas de \hat{h}^2 verificou-se alta herdabilidade para o caráter pelo procedimento Clássico ($\hat{h}^2 = 86,8\%$) e baixa herdabilidade $\hat{h}^2 = 15\%$ pelo método de Bayesiano;

A estimativa de $|d|/|a| = 1,12$ pelo método CIM indicando que o caráter produção de grãos apresenta, em média, dominância completa foi influenciada pelo alto valor de R_G^2 dos QTLs que apresentaram interação alélica de sobredominância, cujos alelos favoráveis são provenientes da linhagem parental menos vigorosa, L20-01. Caso contrário, o caráter poderia apresentar em média dominância parcial, como no método BIM;

O número de QTLs mapeados pelo método CIM depende do limite crítico, “*threshold*”, adotado. Ao adotar nesse trabalho um “*threshold*” com $LOD = 2,5$ para se declarar QTLs, possibilitou que dez QTLs fossem detectados. Caso considerasse, por exemplo, $LOD = 3,0$, o número de QTLs mapeados por este procedimento seria menor;

Com a utilização da abordagem Bayesiana não foi necessário determinar o limiar crítico para declarar a presença de QTLs, visto que o número de QTLs é determinado pelo Fator de Bayes, calculado em uma amostra da distribuição *a posteriori* obtida a partir de métodos

computacionalmente intensivos, como o RJMCMC. Também é possível calcular intervalos de credibilidade $(1 - \alpha)\%$ que garantem que o parâmetro de interesse pertence a esse intervalo;

Adotando-se nesse trabalho o procedimento BIM, o número de QTLs P foi considerado como variável aleatória, cujo valor foi determinado pelo Fator de Bayes, por essa razão foram utilizados os métodos de amostragem MCMC com saltos reversíveis (*Reversible Jump MCMC*). No entanto, a partir de experimentos anteriores o valor de P poderia ser incorporado na análise como informação *a priori*. Nesse caso, o algoritmo Metropolis-Hastings poderia ser utilizado, como sugerido por Satagopan et al. (1996);

Ambos os procedimentos de mapeamento de QTLs tiveram por base apenas o desempenho médio dos genótipos, o que pode ter influenciado na detecção de um número relativamente baixo de QTLs, comparando-se ao trabalho realizado por Bento (2006) com a mesma população, que considerou nove ambientes e mapeou vinte e quatro QTLs associados a produção de grãos;

3 CONCLUSÕES

Os resultados obtidos nesse trabalho possibilitaram as seguintes conclusões:

Através do mapeamento por intervalo composto foram mapeados dez QTLs associados a produção de grãos e cinco QTLs pelo mapeamento por intervalo Bayesiano, concentrados principalmente entre os cromossomos três e cinco, dos quais apenas dois são coincidentes, ($qprod5c = qprod5a^*$) e ($qprod5d = qprod5b^*$), localizados do cromossomo cinco;

Os dez QTLs mapeados pelo método CIM explicam conjuntamente 19,86% da variação genética e 17,09% da variação fenotípica para o caráter produção de grãos. Pelo procedimento BIM os cinco QTLs mapeados explicam 13,06% e 1,93% das variações genética e fenotípica, respectivamente. Em ambos os métodos o grau médio de dominância foi de dominância: completa pelo CIM e parcial pelo BIM;

Os efeitos de sobredominância, cujos alelos favoráveis são provenientes da linhagem parental menos vigorosa, L20-01, poderiam ser explicados ao fato de existirem vários QTLs ligados que não foram detectados. Saturando mais o mapa, evitaria possíveis problemas de ligação entre QTLs e as estimativas de seus parâmetros seriam mais precisas;

O método de mapeamento por intervalo composto evita que um QTL presente em um dado intervalo interfira no mapeamento do intervalo adjacente, incluindo no modelo a informação de marcadores adjacentes ao intervalo considerado sob a forma de cofatores, tornando o método mais informativo e preciso, pois elimina o efeito de QTLs inexistentes. No entanto, a interação entre locos de diferentes alelos é ignorado. Outro problema está no fato de que a eficiência do método depende do nível de saturação do mapa, uma vez que é admitido apenas um QTL por intervalo. Isso ocorre quando a distância entre marcas é relativamente grande, como por exemplo a distância de 54 cM entre os marcadores um (U1177) e dois (B1178) do primeiro cromossomo no mapa genético do milho (Figuras 5, 11 e 15), cujo intervalo poderia conter vários QTLs não detectados;

Um cuidado ao se utilizar o método CIM está em determinar o valor do “*thresh-*

old”, o limite crítico adotado para declarar a presença de um QTL no intervalo analisado. Ao adotar, por exemplo, $LOD = 2$, admite-se que a hipótese alternativa referente à ocorrência de um QTL é $10^2 = 100$ vezes mais provável de ocorrer do que a hipótese nula da não ocorrência. Conseqüentemente, ao adotar um valor como $LOD = 3$, o número de QTLs mapeados será menor;

Utilizando a abordagem Bayesiana não foi necessário determinar o “*threshold*”, para declarar a presença de QTLs, visto que o número de QTLs é determinado pelo Fator de Bayes, calculado em uma amostra da distribuição *a posteriori* obtida a partir de métodos computacionalmente intensivos, como o RJMCMC. A vantagem de se fazer primeiramente a seleção do modelo que melhor se ajusta aos dados fenotípicos, isto é, verificar o número de QTLs associados ao caráter de interesse está na possibilidade de se observar quando um QTL está ou não afetando o caráter em questão;

A vantagem de se utilizar a abordagem Bayesiana está em utilizar informações adicionais (distribuições *a priori*) que conectadas às informações dos dados observados (verossimilhança) são atualizadas através do Teorema de Bayes. Essa atualização sobre o parâmetro de interesse θ é a distribuição *a posteriori*. Quanto menos informativa for *a priori* mais alta será a variabilidade e, conseqüentemente, maior será a informação fornecida pelos dados. Dessa forma, as estimativas das estatísticas Clássica e Bayesiana estarão mais próximas;

Uma fragilidade verificada na metodologia Bayesiana de mapeamento por intervalo foi com relação a especificação das *prioris*. Ao efetuar o mapeamento de QTLs, através da análise de sensibilidade onde foram alteradas as densidades *a priori* nos parâmetros ou nas distribuições, foi verificado que as análises foram significativamente afetadas. Deve, portanto, haver cautela ao especificar as densidades *a priori*;

Outro cuidado ao adotar o procedimento BIM está no tamanho da amostra *posteriori* resultante do número de iterações dos métodos MCMC para que os resultados se tornem estáveis. Foi verificada instabilidade nas estimativas dos resumos *a posteriori* em amostras geradas de tamanho 1000;

Um procedimento interessante seria utilizar o mapeamento por intervalo Bayesiano considerando múltiplos ambientes. Além disso, adicionar, se possível, informações *a priori* dos cofatores associados às marcas, bem como a presença de interação genótipo \times ambiente que, a partir de tais informações incorporadas como informações *a priori*, poderão aumentar significativamente a precisão da análise estatística.

REFERÊNCIAS

- BARBIN, D. **Componentes de variância: teoria e aplicações**. 2ed. Piracicaba: FEALQ, 1993. 120p.
- BARBIN, D. **Planejamento e análise estatística de experimentos agronômicos**. Arapongas: Midas, 2003. 194p.
- BEARZOTI, E. Mapeamento de QTL In: PINHEIRO, J. B. **Análise de QTL no melhoramento de plantas: Segunda jornada em Genética e melhoramento de plantas**. Goiânia: FUNAPE, 2000, p. 63-209
- BENTO, D.A.V. **Mapeamento de QTLs para produção de grãos e seus componentes em uma população de milho tropical**. 2006. 135 p. Tese (Doutorado em Genética e Melhoramento de Plantas) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 2006.
- BOX, G.E.; TIAO, G.C. **Bayesian inference in statistical analysis**. New York: Wiley, 1992. 588 p.
- BOLFARINE, H. SANDOVAL, M.C. **Introdução à Inferência Estatística** Rio de Janeiro: Sociedade Brasileira de Matemática, 2001. 125 p.
- BUSSAB, W.O.; MORETIN, P.A. **Estatística básica**. 5. ed. São Paulo: Saraiva, 2003. 526 p.
- COELHO, A.S.G. ; SILVA, H.D. **Métodos biométricos aplicados à análise de QTLs**. In: REUNIÃO ANUAL DA REGIÃO BRASILEIRA DA SOCIEDADE INTERNACIONAL DE BIOMETRIA, 50. E SIMPÓSIO DE ESTATÍSTICA APLICADA À EXPERIMENTAÇÃO AGRONÔMICA, 11., Londrina. **Anais e Resumos**. Londrina: UEL, 2005, 1 CD-ROM .
- EHLERS, R.S., **Introdução à inferência Bayesiana** Curitiba: Universidade Federal do Paraná. Departamento de Estatística, 2003. 55 p.
- FALCONER, D.S.; MACKAY, T.F.C. **Introduction to quantitative genetics**. 4th ed. London: Longman Scientific and Tecnical, 1996. 464 p.
- GAFFNEY, P.J. **An efficient reversible jump markov chain monte carlo approach to detect multiple loci and their effects in inbred crosses**. 2001. 174 p. Tese (Doctor of Philosophy in Statistics) - University Of Wisconsin, Madison, 2001.

GARCIA, A.F. **Mapeamento de QTLs para caracteres de importância agronômica em duas populações F_2 de milho tropical**. 2003. 110 p. Dissertação (Mestrado em Genética Vegetal e Melhoramento) - Universidade Estadual de Campinas, Campinas, 2003.

GELFAND, A.E.; SMITH, A.F.M. Sampling-based approaches to calculating marginal densities. **Journal of American Statistical Association**, v. 85, p. 398-410, 1990.

GEMAN, S.; GEMAN, D. Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 6, p. 169-193, 1984.

GREEN, P.J. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. **Biometrika** v. 82, p. 711-732; 1995.

GRIFFITHS, A. J.F.; MILLER, J.H.; SUZUKI, D.T.; LEWONTIN, R.C.; GELBART, W.M. **Introdução à genética**. 7. ed. Rio de Janeiro: Guanabara Koogan, 2002. 470p.

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

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

JANSEN, R.C.; STAM, P. Multiple trait analysis of genetic mapping for quantitative trait loci. **Genetics**, Austin, v. 140, p. 1111-1127, 1995.

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

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

LEANDRO, R.A. **Introdução à estatística Bayesiana**. In: REUNIÃO ANUAL DA REGIÃO BRASILEIRA DA SOCIEDADE INTERNACIONAL DE BIOMETRIA, 46.; SIMPÓSIO DE ESTATÍSTICA APLICADA À EXPERIMENTAÇÃO AGRONÔMICA, 9., 2001, Piracicaba. **Minicurso** Piracicaba: ESALQ/USP, 2001. 51p.

- LIMA, L.A.L. **Mapeamento de QTL para produção de grãos e caracteres de planta em milho tropical utilizando marcadores microssatélites**. 2006. 110 p. Tese (Doutorado em Genética Vegetal e Melhoramento) - Universidade Estadual de Campinas, Campinas, 2006.
- LIU, B.H. **Statistical Genomics: Linkage, Mapping, and QTL Analysis**. Boca Raton, FL: CRC Press, 1998. 611 p.
- LYNCH, M.; WALSH, B. **Genetics and analysis of quantitative traits.**, Massachusetts: Sinauer Sunderland, 1998, 980 p.
- MANGOLIN, C.A. **Mapeamento de QTLs associados com o teor de óleo em linhagens de milho (*Zea mays L.*)**. 2002. 123 p. Tese (Doutorado em Genética Vegetal e Melhoramento) - Universidade Estadual de Campinas, Campinas, 2002.
- MANGOLIN, C.A.; SOUZA Jr, C. L; GARCIA, A. F.; GARCIA, A. A. F.; SIBOV, S.T; SOUZA, A. P. Mapping QTLs for kernel oil in a tropical maize population. **Euphytica**, Wageningen, v. 137, n. 2, p. 251-259, 2004.
- METROPOLIS, N.; ROSENBLUTH, M.; TELLER, E. Equations of state calculations by fast computing machines. **Journal of Chemical Physics**, v. 21, p. 1087-1092, 1953.
- O'HAGAN, A.; LUCE, B.R.; FRYBACK, D.G. **A primer on Bayesian statistics in health economics and outcomes research**. Pennsylvania: Medtap International, 2003, 72 p.
- PAULINO, C.D.; TURKMAN, M.A.; MURTEIRA, B., **Estatística Bayesiana** Lisboa: Fundação Calouste Gulbenkian, 2003. 446 p.
- RAFTERY, A. E. Bayesian model selection in social research. **Sociological Methodology**, v. 25, p. 111-163, 1995.
- RAGONHA, F. **Estimação da frequência de recombinação em fase parental desconhecida: uma abordagem Bayesiana**. 2005. 130 p. Dissertação (Mestrado em Estatística e Experimentação Agronômica) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 2005.
- RAMALHO, M.A.P.; SANTOS, J.B.; PINTO, C.A.B.P., **Genética na agropecuária** 3.ed rev. Lavras: UFLA, 2004. 472 p.

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**, p. 805-816, 1996.

SCHUSTER, I.; CRUZ, C.D. **Estatística genômica aplicada a populações derivadas de cruzamentos controlados**. Viçosa: UFV, 2004. 568 p.

SILLANPÄÄ, M. J.; ARJAS, E. Bayesian mapping of multiple quantitative trait loci from incomplete inbred line cross data. **Genetics** v. 148, p. 1373-1388, 1998.

SILVA, H.D. **Aspectos biométricos da detecção de QTLs (Quantitative Trait Loci) em espécies cultivadas**. 2001. 130 p. Tese (Doutorado em Estatística e Experimentação Agrônômica) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, 2001.

SOUZA, A.D.P. **Fundamentos de inferência Bayesiana e aplicações**. In: SEMANA DOS CURSOS DE MATEMÁTICA, ESTATÍSTICA E CIÊNCIA DA COMPUTAÇÃO, 2003, Presidente Prudente. **Minicurso** Presidente Prudente: UNESP 2003. 35 p.

SOUZA, A.D.P.; TACHIBANA, V. M. **Introdução aos processos estocásticos**. Presidente Prudente: Universidade Estadual Paulista. Departamento de Matemática, 2001, 71 p.

YANDELL, B.S. **A Bayesian model selection for quantitative trait loci with Markov chain Monte Carlo in experimental crosses**. Disponível em: <www.stat.wisc.edu/yandell/statgen>. Acesso em: 10 de maio de 2005.

YANDELL, B.S.; GAFFNEY, P.J.; SATAGOPAN, J.M.; JIN, C.; WU H. **Bayesian interval mapping**. Disponível em: <www.stat.wisc.edu/yandell/statgen>. Acesso em: 10 de ago. de 2005.

SORENSEN, D.; GIANOLA, D. **Statistics for biology and health likelihood, Bayesian, and MCMC methods in quantitative genetics**. USA: Springer-Verlag, 2002. 479 p.

WINDOWS QTL CARTOGRAPHER. **Windows QTL Cartographer 2.5 User Manual**. North Carolina State University, 2005. Disponível em: <<http://www.cs.wisc.edu/yandell/ql/software/Bmapqtl/Bmapqtl.pdf>>. Acesso em: 2 de abr. de 2005, p. 67-73.

ZENG, Z.B. Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci. **Procedures of the National Academic Science USA** , v. 90, p. 10972-10976, 1993.

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

ZENG, Z.B. **Mapping QTL**. In: REUNIÃO ANUAL DA REGIÃO BRASILEIRA DA SOCIEDADE INTERNACIONAL DE BIOMETRIA, 46.; SIMPÓSIO DE ESTATÍSTICA APLICADA À EXPERIMENTAÇÃO AGRONÔMICA, 9., 2001, Piracicaba. **Minicurso** Piracicaba: ESALQ/USP, 2001. 57p.

ZENG, Z.B. Modeling epistasis of quantitative trait loci using Cockerham's model. **Genetics**, v. 160, p. 1243-1261, 2002.

ANEXOS

ANEXO A - Programa em *R* utilizado para calcular o Fator de Bayes, para determinar o número de QTLs pelo procedimento BIM.

```
rm(list=ls(all=TRUE))
# Fator de Bayes para Produção de Grãos
dados<-read.table("e:dadFBayesPG.txt",header=T) # Priori
dados0<-NULL
dados1<-NULL
dados2<-NULL
dados3<-NULL
dados4<-NULL
dados5<-NULL
dados6<-NULL
dados7<-NULL
dados8<-NULL
dados9<-NULL
dados10<-NULL
dados11<-NULL
dados12<-NULL
dados13<-NULL
for(i in 1:14)
{
dados0[i] <- dados[1,3]/dados[i,3]
dados1[i] <- dados[2,3]/dados[i,3]
dados2[i] <- dados[3,3]/dados[i,3]
dados3[i] <- dados[4,3]/dados[i,3]
dados4[i] <- dados[5,3]/dados[i,3]
dados5[i] <- dados[6,3]/dados[i,3]
dados6[i] <- dados[7,3]/dados[i,3]
dados7[i] <- dados[8,3]/dados[i,3]
dados8[i] <- dados[9,3]/dados[i,3]
dados9[i] <- dados[10,3]/dados[i,3]
dados10[i] <- dados[11,3]/dados[i,3]
dados11[i] <- dados[12,3]/dados[i,3]
dados12[i] <- dados[13,3]/dados[i,3]
dados13[i] <- dados[14,3]/dados[i,3]
}
# posteriori
dadoso1<-NULL
dados11<-NULL
dados21<-NULL
dados31<-NULL
dados41<-NULL
dados51<-NULL
dados61<-NULL
dados71<-NULL
dados81<-NULL
dados91<-NULL
dados101<-NULL
```

```

dados111<-NULL
dados121<-NULL
dados131<-NULL
for(i in 1:14)
{
  dadoso1[i] <- dados[1,2]/dados[i,2]
  dados11[i] <- dados[2,2]/dados[i,2]
  dados21[i] <- dados[3,2]/dados[i,2]
  dados31[i] <- dados[4,2]/dados[i,2]
  dados41[i] <- dados[5,2]/dados[i,2]
  dados51[i] <- dados[6,2]/dados[i,2]
  dados61[i] <- dados[7,2]/dados[i,2]
  dados71[i] <- dados[8,2]/dados[i,2]
  dados81[i] <- dados[9,2]/dados[i,2]
  dados91[i] <- dados[10,2]/dados[i,2]
  dados101[i] <- dados[11,2]/dados[i,2]
  dados111[i] <- dados[12,2]/dados[i,2]
  dados121[i] <- dados[13,2]/dados[i,2]
  dados131[i] <- dados[14,2]/dados[i,2]
}
priori<-cbind(dados0,dados1,dados2,dados3,dados4,dados5,dados6,dados7,dados8,dados9,dados10,dados11,dados12,dados13)
post<-cbind(dadoso1,dados11,dados21,dados31,dados41,dados51,dados61,dados71,dados81,dados91,dados101,dados111,dados121,dados131)
priori post
M0<-NULL M1<-NULL M2<-NULL M3<-NULL M4<-NULL M5<-NULL M6<-NULL
M7<-NULL M8<-NULL M9<-NULL M10<-NULL M11<-NULL M12<-NULL M13<-NULL
for(i in 1:14)
{
  M0[i]<-post[i,1]/priori[i,1]
  M1[i]<-post[i,2]/priori[i,2]
  M2[i]<-post[i,3]/priori[i,3]
  M3[i]<-post[i,4]/priori[i,4]
  M4[i]<-post[i,5]/priori[i,5]
  M5[i]<-post[i,6]/priori[i,6]
  M6[i]<-post[i,7]/priori[i,7]
  M7[i]<-post[i,8]/priori[i,8]
  M8[i]<-post[i,9]/priori[i,9]
  M9[i]<-post[i,10]/priori[i,10]
  M10[i]<-post[i,11]/priori[i,11]
  M11[i]<-post[i,12]/priori[i,12]
  M12[i]<-post[i,13]/priori[i,13]
  M13[i]<-post[i,14]/priori[i,14]
}
fbayes<-cbind(M0,M1,M2,M3,M4,M5,M6,M7,M8,M9,M10,M11,M12,M13)#B(Mi+1, Mi)
rownames(fbayes)<-c("M0", "M1", "M2", "M3", "M4", "M5", "M6", "M7", "M8", "M9", "M10", "M11", "M12", "M13")
fbayes

```

ANEXO B - Programa em *R* utilizado para obter os gráficos das seqüências MCMC e densidades *a posteriori* marginais $\pi(\mathbf{a}_j|\mathbf{y})$, $\pi(\mathbf{d}_j|\mathbf{y})$ e $\pi(\lambda_j|\mathbf{y})$ dos efeitos de aditividade (a_j), de dominância (d_j) e posição (λ_j) do j -ésimo QTL no cromossomo ($j = 1, \dots, 5$).

```
#####
#Conjunto de dados Produção de Grãos: # média 43,9059; s2=71,2898 e
#400000 iterações
#####
rm(list=ls(all=TRUE))
dados<-read.table("e:dad4PG.txt",header=T,na.strings=".")
attach(dados) dim(dados) #dad5: modelo com 5 QTLs - Análise
Exploratória de M5: dad5<-dados[which(nqtl==5),] dim(dad5)
summary(dad5[,5:13]) dad61<-dados[which(nqtl==5 & iqtl==1),]
dad62<-dados[which(nqtl==5 & iqtl==2),]
dad63<-dados[which(nqtl==5&iqtl==3),]
dad64<-dados[which(nqtl==5&iqtl==4),]
dad65<-dados[which(nqtl==5&iqtl==5),] dim(dad61)dim(dad62)dim(dad63)
dim(dad64) dim(dad65)
#####
# Efeitos genéticos e localização dos QTLs detectados pelo método
#Bayesiano
#####
# QTL1: qprod1a* par(mfrow=c(3,2))
ts.plot(dad61[,10],main=list("QTL1: qprod1a*", cex=1.2,
col="black",font=2),xlab='Seqüência
MCMC',ylab=expression(paste(italic(a[1]))) #aditivo abline
h=mean(dad61[,10]), col="red", lwd=2)
hist(dad61[,10],prob=T,main=list("Posteriori", cex=1.2, col="black",
font=2), xlab=expression(paste("Ef. aditivo, ",italic(a[1])))
,ylab=expression(paste("P(", italic(a[1]),"|",bold(y),")")))#aditivo
a1<-density(dad61[,10]) lines(a1, col="blue", lwd=2)
ts.plot(dad61[,11], xlab='Seqüência
MCMC',ylab=expression(paste(italic(d[1]))) # dom
abline(h=mean(dad61[,11]), col="red", lwd=2)
hist(dad61[,11],prob=T,main=list(" ", cex=1.2, col="black", font=2),
xlab=expression(paste("Ef. dominância, ",italic(d[1])))
,ylab=expression(paste("P(", italic(d[1]),"|",bold(y),")")) ) # dom
d1<-density(dad61[,11]) lines(d1, col="blue", lwd=2)
ts.plot(dad61[,12], xlab='Seqüência
MCMC',ylab=expression(paste(lambda[1])) ) # locus
abline(h=mean(dad61[,12]), col="red", lwd=2)
hist(dad61[,12],prob=T,main=list(" ", cex=1.2, col="black", font=2),
xlab=expression(paste("Posição, ",lambda[1]))
,ylab=expression(paste("P(", lambda[1],"|",bold(y),")")) ) # locus
l1<-density(dad61[,12]) lines(l1, col="blue", lwd=2)
#####
# QTL2: qprod3a* par(mfrow=c(3,2)) ts.plot(dad62[
,10],main=list("QTL2: qprod3a*", cex=1.2, col="black", font=2),
xlab='Seqüência MCMC',ylab=expression(paste(italic(a[2]))) ) #
```

```

aditivo abline (h=mean(dad62[ ,10]), col="red", lwd=2)
hist(dad62[,10],prob=T,main=list("Posteriori", cex=1.2, col="black",
font=2), xlab=expression(paste("Ef. aditivo", ",italic(a[2]")))
,ylab=expression(paste("P(", italic(a[2]),"|",bold(y),")")))#aditivo
a2<-density(dad62[ ,10]) lines(a2, col="blue", lwd=2)
ts.plot(dad62[,11], xlab='Seqüência
MCMC',ylab=expression(paste(italic(d[2])))) # dom
abline(h=mean(dad62[ ,11]), col="red", lwd=2)
hist(dad62[,11],prob=T,main=list(" ", cex=1.2, col="black", font=2),
xlab=expression(paste("Ef. dominância", ",italic(d[2]")))
,ylab=expression(paste("P(", italic(d[2]),"|",bold(y),")")) ) # dom
d2<-density(dad62[ ,11]) lines(d2, col="blue", lwd=2)
ts.plot(dad62[,12], xlab='Seqüência
MCMC',ylab=expression(paste(lambda[2])) ) # locus
abline(h=mean(dad62[ ,12]), col="red", lwd=2) hist(dad62[
,12],prob=T,main=list(" ", cex=1.2, col="black", font=2),
xlab=expression(paste("Posição", ",lambda[2]")))
,ylab=expression(paste("P(", lambda[2],"|",bold(y),")")) # locus
l2<-density(dad62[ ,12]) lines(l2, col="blue", lwd=2)
#####
# QTL3: qprod5a* par(mfrow=c(3,2)) ts.plot(dad63[
,10],main=list("QTL3: qprod5a*", cex=1.2, col="black", font=2),
xlab='Seqüência MCMC',ylab=expression(paste(italic(a[3]))) )#aditivo
abline (h=mean(dad63[ ,10]), col="red", lwd=2)
hist(dad63[,10],prob=T,main=list("Posteriori", cex=1.2, col="black",
font=2), xlab=expression(paste("Ef. aditivo", ",italic(a[3]")))
,ylab=expression(paste("P(", italic(a[3]),"|",bold(y),")")))#aditivo
a3<-density(dad63[ ,10]) lines(a3, col="blue", lwd=2)
ts.plot(dad63[,11], xlab='Seqüência
MCMC',ylab=expression(paste(italic(d[3])))) #dom
abline(h=mean(dad63[ ,11]), col="red", lwd=2)
hist(dad63[,11],prob=T,main=list(" ", cex=1.2, col="black", font=2),
xlab=expression(paste("Ef. dominância", ",italic(d[3]")))
,ylab=expression(paste("P(", italic(d[3]),"|",bold(y),")")) ) # dom
d3<-density(dad63[ ,11]) lines(d3, col="blue", lwd=2)
ts.plot(dad63[,12], xlab='Seqüência
MCMC',ylab=expression(paste(lambda[3])) ) # locus abline
(h=mean(dad63[ ,12]), col="red", lwd=2) hist(dad63[
,12],prob=T,main=list(" ", cex=1.2, col="black", font=2),
xlab=expression(paste("Posição", ",lambda[3]")))
,ylab=expression(paste("P(", lambda[3],"|",bold(y),")")) # locus
l3<-density(dad63[ ,12]) lines(l3, col="blue", lwd=2)
#####
# QTL4: qprod5b* par(mfrow=c(3,2)) ts.plot(dad64[
,10],main=list("QTL4: qprod5b*", cex=1.2, col="black", font=2),
xlab='Seqüência MCMC',ylab=expression(paste(italic(a[4]))) )#aditivo
abline (h=mean(dad64[ ,10]), col="red", lwd=2)
hist(dad64[,10],prob=T,main=list("Posteriori", cex=1.2, col="black",
font=2), xlab=expression(paste("Ef. aditivo", ",italic(a[4])))

```

```

,ylab=expression(paste("P(", italic(a[4]),"|",bold(y),")"))#aditivo
a4<-density(dad64[,10]) lines(a4, col="blue", lwd=2)
ts.plot(dad64[,11],xlab='Seqüência
MCMC',ylab=expression(paste(italic(d[4]))) # dom
abline(h=mean(dad64[,11]), col="red", lwd=2)
hist(dad64[,11],prob=T,main=list(" ", cex=1.2, col="black", font=2),
xlab=expression(paste("Ef. dominância, ",italic(d[4])))
,ylab=expression(paste("P(", italic(d[4]),"|",bold(y),")")) ) # dom
d4<-density(dad64[,11]) lines(d4, col="blue", lwd=2)
ts.plot(dad64[,12], xlab='Seqüência
MCMC',ylab=expression(paste(lambda[4])) ) # locus
abline(h=mean(dad64[,12]), col="red", lwd=2)
hist(dad64[,12],prob=T,main=list(" ", cex=1.2, col="black", font=2),
xlab=expression(paste("Posição, ",lambda[4]))
,ylab=expression(paste("P(", lambda[4],"|",bold(y),")")) # locus
l4<-density(dad64[,12]) lines(l4, col="blue", lwd=2)
#####
# QTL5: qprod5c* par(mfrow=c(3,2)) ts.plot(dad65[
,10],main=list("QTL5: qprod5c*", cex=1.2, col="black", font=2),
xlab='Seqüência MCMC',ylab=expression(paste(italic(a[5]))) #aditivo
abline (h=mean(dad65[,10]), col="red", lwd=2)
hist(dad65[,10],prob=T,main=list("Posteriori", cex=1.2, col="black",
font=2), xlab=expression(paste("Ef. aditivo, ",italic(a[5])))
,ylab=expression(paste("P(", italic(a[5]),"|",bold(y),")"))#aditivo
a5<-density(dad65[,10]) lines(a5, col="blue", lwd=2)
ts.plot(dad65[,11], xlab='Seqüência
MCMC',ylab=expression(paste(italic(d[5]))) # dom
abline(h=mean(dad65[,11]), col="red", lwd=2)
hist(dad65[,11],prob=T,main=list(" ", cex=1.2, col="black", font=2),
xlab=expression(paste("Ef. dominância, ",italic(d[5])))
,ylab=expression(paste("P(", italic(d[5]),"|",bold(y),")")) ) # dom
d5<-density(dad65[,11]) lines(d5, col="blue", lwd=2)
ts.plot(dad65[,12], xlab='Seqüência
MCMC',ylab=expression(paste(lambda[5])) ) # locus
abline(h=mean(dad65[,12]), col="red", lwd=2)
hist(dad65[,12],prob=T,main=list(" ", cex=1.2, col="black", font=2),
xlab=expression(paste("Posição, ",lambda[5]))
,ylab=expression(paste("P(", lambda[5],"|",bold(y),")")) # locus
l5<-density(dad65[,12]) lines(l5, col="blue", lwd=2)

```

ANEXO C - Programa em *R* utilizado para obter os Intervalos de Credibilidade para \mathbf{a}_j , \mathbf{d}_j e λ_j .

```

a1<-quantile(dad61[,10], c(0.025, 0.975)) #aditivo QTL 1
d1<-quantile(dad61[,11], c(0.025, 0.975)) # dom
l1<-quantile(dad61[,12], c(0.025, 0.975)) # locus
a2<-quantile(dad62[,10], c(0.025, 0.975)) # aditivo QTL 2
d2<-quantile(dad62[,11], c(0.025, 0.975)) # dom
l2<-quantile(dad62[,12], c(0.025, 0.975)) # locus
a3<-quantile(dad63[,10], c(0.025, 0.975)) # aditivo QTL 3
d3<-quantile(dad63[,11], c(0.025, 0.975)) # dom
l3<-quantile(dad63[,12], c(0.025, 0.975)) # locus
a4<-quantile(dad64[,10], c(0.025,0.975))#aditivo QTL 4
d4<-quantile(dad64[,11], c(0.025, 0.975)) # dom
l4<-quantile(dad64[,12], c(0.025, 0.975)) # locus
a5<-quantile(dad66[,10], c(0.025, 0.975)) # aditivo QTL 5
d5<-quantile(dad66[,11], c(0.025, 0.975)) # dom
l5<-quantile(dad66[,12], c(0.025, 0.975)) # locus
IC_1<-cbind(a1,d1,l1) IC_2<-cbind(a2, d2, l2) IC_3<-cbind(a3, d3,
l3) IC_4<-cbind(a4, d4, l4) IC_5<-cbind(a5, d5, l5)
IC<-rbind(IC_1,IC_2,IC_3,IC_4,IC_5)
IC

```

Livros Grátis

(<http://www.livrosgratis.com.br>)

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

[Baixar livros de Literatura](#)
[Baixar livros de Literatura de Cordel](#)
[Baixar livros de Literatura Infantil](#)
[Baixar livros de Matemática](#)
[Baixar livros de Medicina](#)
[Baixar livros de Medicina Veterinária](#)
[Baixar livros de Meio Ambiente](#)
[Baixar livros de Meteorologia](#)
[Baixar Monografias e TCC](#)
[Baixar livros Multidisciplinar](#)
[Baixar livros de Música](#)
[Baixar livros de Psicologia](#)
[Baixar livros de Química](#)
[Baixar livros de Saúde Coletiva](#)
[Baixar livros de Serviço Social](#)
[Baixar livros de Sociologia](#)
[Baixar livros de Teologia](#)
[Baixar livros de Trabalho](#)
[Baixar livros de Turismo](#)