

Modelos Aleatórios na Estimação da Localização de QTLs em Famílias de Meios-Irmãos¹

Marcos Vinicius G. Barbosa da Silva^{2,3}, Mário Luiz Martinez^{2,3}, Robledo de Almeida Torres^{3,4}, Paulo Sávio Lopes^{3,4}, Ricardo Frederico Euclides^{3,4}, Carmen Silva Pereira^{3,4}, Marco Antônio Machado², Wagner Arbex⁵

RESUMO - O procedimento de mapeamento por intervalo baseado em modelo aleatório foi aplicado para estudar sua robustez e propriedades no mapeamento de dois QTLs, em populações constituídas de famílias de meios-irmãos. Sob o modelo aleatório, a localização e os componentes de variância foram estimados usando o método da máxima verossimilhança. A estimação dos parâmetros foi baseada na metodologia de pares de irmãos. As proporções de genes idênticos por descendência (IBD) dos QTLs foram estimadas a partir das proporções IBD de dois marcadores flanqueadores. As estimativas dos parâmetros dos QTLs (localizações e componentes de variância) e do poder de detecção, em uma característica com $h^2 = 0,25$, foram obtidas usando dados simulados, variando-se o número de famílias, o tamanho das famílias e a proporção da variância genética devida aos QTLs e à posição dos QTLs, localizados no mesmo intervalo, em intervalos adjacentes e não-adjacentes. Os fatores que mais influenciaram as estimativas dos parâmetros foram as proporções da variância devida aos QTLs, ao número e ao tamanho das famílias. As estimativas mais viesadas dos componentes dos QTLs e poligênico foram obtidas com a análise dos registros de somente dez famílias. Com número suficiente de famílias e de indivíduos nas famílias e altas proporções de variância genética devida aos QTLs, o modelo aleatório pode detectar QTL com alto poder, apresentando estimativas das posições com boa acurácia, se não existirem dois QTLs no mesmo intervalo. Para o mapeamento fino e estimativas apropriadas da variância dos QTLs, métodos mais sofisticados devem ser utilizados.

Palavras-chave: mapeamento por intervalo, marcador genético, método dos pares de irmãos, modelo aleatório, QTL

Random Model Approach for Estimation of QTL Location in Half-Sibs Families

ABSTRACT - An interval mapping procedure based on the random model approach was applied to investigate its robustness and properties for QTL mapping in populations with prevailing half-sib structures. Under random model, QTL location and variance components were estimated using maximum likelihood procedures. The estimation of parameters was based on sib-pair approach. The proportions of genes identical-by-descent (IBD) at the two QTLs were estimated from the IBD at two flanking marker loci. Estimates for QTLs parameters (locations and variance components) and power of detection, in a trait with $h^2 = 0.25$, were obtained using simulated data, varying the number of families, number of half-sibs by families, proportion of QTL variance and QTLs positions. These QTLs were positioned at the same interval, at adjacent intervals and at no adjacent intervals. The most important factors influencing the estimates of QTLs parameters and power were the proportion of variance due to QTLs the number and size of the half-sibs families. The most biased variance components estimations were obtained when sample size was ten. Given a sufficient number of families and high proportions of genetic variance due to QTLs, the random model approach can detect a QTLs with high power and provides accurate estimates of the QTLs position if two QTLs are not present in the same interval. For fine QTL mapping and proper estimation of QTLs variance, more sophisticated methods are required.

Key Words: genetic marker, interval mapping, QTL, random model, sib-pair approach

Introdução

A identificação e o mapeamento de genes que influenciam as características quantitativas, denominadas *quantitative trait loci* ou simplesmente QTLs, têm recebido muita atenção nos últimos anos (Kadarmideen & Dekeers, 1999), principalmente com

o desenvolvimento de mapas de ligação com grande número de marcadores moleculares. O principal objetivo, nesse caso, é acelerar o progresso genético com o uso dos QTLs identificados.

Para identificação desses locos, geralmente, são utilizados dados provenientes de cruzamentos entre linhagens endogâmicas (ou raças), que incluem, por

¹ Parte da tese de Doutorado do primeiro autor apresentada à UFV.

² Pesquisador da Embrapa Gado de Leite - Rua Eugênio do Nascimento, 610 -36038-330, Juiz de Fora, MG E-mail: marcos@cnpqgl.embrapa.br

³ Bolsista do CNPq.

⁴ Professor do DZO/ UFV.

⁵ Técnico de nível superior da Embrapa Gado de Leite.

exemplo, o retrocruzamento e o F_2 , ou de populações chamadas segregantes ou exogâmicas (Xu & Atchley, 1995), como ocorre na maioria das espécies de animais domésticos e em árvores frutíferas perenes.

A análise de registros de cruzamentos entre linhagens endogâmicas apresenta algumas vantagens, como o conhecimento dos genótipos parentais, da fase de ligação e do número de alelos do suposto QTL. Esse tipo de dado também pode ser considerado como sendo uma única grande família, porque todos os indivíduos compartilham o mesmo genótipo parental, o que resulta em estimação direta dos efeitos de substituição e de dominância do QTL (Zeng, 1994). Todavia, muitas vezes, a implementação desses cruzamentos é impraticável e onerosa.

Os dados de campo, por sua vez, podem ser obtidos de modo mais fácil e barato; entretanto, apresentam como desvantagens o fato de os genótipos parentais não serem conhecidos, a existência de muitas famílias distintas com probabilidades condicionais do genótipo do QTL, dado o genótipo do marcador, diferentes, as fases de ligação e o número de alelos do QTL são desconhecidos e o tamanho da família, em geral, é reduzido, o que impede a obtenção de estimativas mais precisas.

Desse modo, de acordo com Amos & Elston (1989), foram desenvolvidos métodos robustos com base em modelos aleatórios (Haseman & Elston, 1972; Fulker & Cardon, 1994), que não requerem a especificação de modelos genéticos para a detecção da ligação.

Os modelos aleatórios são baseados no compartilhamento de diferentes proporções de genes idênticos por descendência (IBD) entre dois parentes, isto é, quanto maior a proporção de genes IBD, maior será a similaridade fenotípica (ou covariância) entre dois indivíduos aparentados. A covariância entre eles é composta de um componente devido ao efeito poligênico e outro componente decorrente do efeito do QTL. O componente poligênico é dado pela soma dos efeitos de todos os locos que influenciam a característica, com exceção do suposto QTL. Admite-se, portanto, que ele seja a proporção média de alelos IBD compartilhados por dois indivíduos, ou seja igual ao numerador do coeficiente de parentesco de Wright, isto é, $1/2$ para irmãos completos e $1/4$ para meios-irmãos.

De acordo com Martinez et al. (1999), para o mesmo tipo de parentesco, as proporções de genes IBD em relação ao QTL podem ser diferentes, dado

que os indivíduos podem compartilhar, ou não, os mesmos alelos. Nesse caso, dois irmãos que compartilham um mesmo alelo do QTL serão geneticamente mais semelhantes do que outros dois irmãos que não compartilham nenhum. Obviamente, como não se pode observar diretamente os alelos do QTL, essa proporção é tomada a partir dos genótipos observados dos marcadores.

Martinez & Vukasinovic (2000) estenderam o procedimento proposto por Xu & Atchley (1995) para situações em que há prevalência de famílias de meias-irmãs, como em gado de leite, e mostraram que QTLs com efeitos relativamente grandes podem ser detectados com alto poder e localizados com bastante acurácia, especialmente se grande número de famílias e marcadores polimórficos são usados.

Objetivou-se com este estudo testar a robustez do modelo aleatório no mapeamento de dois QTLs em famílias de meios-irmãos, posicionados no mesmo intervalo, em intervalos adjacentes e não-adjacentes, com diferentes tamanhos e números de famílias e variâncias dos QTLs, em característica com herdabilidade de 0,25.

Material e Métodos

Estimação da proporção de genes IBD em famílias de meios-irmãos

De acordo com Martinez et al. (1999), se os marcadores são totalmente informativos, a proporção de alelos IBD (p_i) compartilhados por dois irmãos completos, em determinado loco, pode ser 0, $1/2$ ou 1, se eles compartilham zero, um ou dois alelos parentais, respectivamente. Em famílias de meios-irmãos, a proporção de alelos IBD em um loco pode somente considerar os valores 0 ou $1/2$, desde que esses indivíduos só possuam um pai em comum.

Como os valores de p_i normalmente não são conhecidos, torna-se necessário estimá-los, obtendo-se as probabilidades condicionais sobre a proporção de IBD, a partir dos tipos de acasalamento entre os pais e os pares de genótipos resultantes. Assim, para qualquer loco do marcador e par de meios-irmãos:

$$\pi_a = f_a / 2$$

em que f_a é a probabilidade de que o par de irmãos tenha um alelo IBD em um loco ao acaso, condicional sobre a informação disponível para o par de irmãos e genótipos parentais (Martinez et al., 1999).

As proporções de alelos IBD nos locos dos marcadores são usadas para calcular a proporção de alelos IBD no loco do QTL, porque duas progênes que recebem os mesmos alelos do marcador, provavelmente, recebem o mesmo alelo do QTL.

Fulker & Cardon (1994) propuseram, então, usar π_1 e π_2 para estimar o número médio de alelos IBD compartilhados pelos irmãos no loco do QTL, π_q , supondo-se vários valores para a taxa de recombinação no intervalo entre marcadores flanqueadores. Assim, a equação para obtenção da estimativa de π_q é:

$$\pi_q = E(\pi_q | \pi_1, \pi_2) = \alpha + \beta_1 \pi_1 + \beta_2 \pi_2$$

em que π_1 e π_2 são os valores IBD para os dois marcadores flanqueadores.

Os β 's são obtidos por meio da seguinte equação normal:

$$\begin{pmatrix} \text{Cov}(\pi_1, \pi_q) \\ \text{Cov}(\pi_2, \pi_q) \end{pmatrix} = \begin{pmatrix} V(\pi_1) & \text{Cov}(\pi_1, \pi_2) \\ \text{Cov}(\pi_1, \pi_2) & V(\pi_2) \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$$

Definindo θ_1 , θ_2 e θ_{12} como as frações de recombinação entre o suposto QTL e o marcador situado à sua esquerda, entre o suposto QTL e o marcador à sua direita e entre os dois marcadores flanqueadores, respectivamente, e substituindo-se todos os valores de π por $1/4$, todas as variâncias ($V(\pi_i)$) por $1/16$ e todas as covariâncias ($\text{Cov}(\pi_i, \pi_j)$) por $(1 - 2\theta_{ij})^2/16$, e resolvendo-se a equação normal, as estimativas dos β 's, segundo Martinez & Vukasinovic (2000), seriam dadas por:

$$\beta_{\pi_2} = \frac{[(1 - 2\theta_2)^2 - (1 - 2\theta_1)^2(1 - 2\theta_{12})^2]}{[1 - (1 - 2\theta_{12})^4]}$$

$$\beta_{\pi_1} = \frac{[(1 - 2\theta_2)^2 - (1 - 2\theta_1)^2(1 - 2\theta_{12})^2]}{[1 - (1 - 2\theta_{12})^4]} e$$

$$\alpha = \frac{(1 - \beta_{\pi_1} - \beta_{\pi_2})}{4}$$

Procedimento de mapeamento por intervalo sob modelo aleatório

A forma geral do modelo definido por Goldgar (1990) é:

$$y_{ij} = \mu + g_{ij} + a_{ij} + e_{ij}$$

em que y_{ij} = característica fenotípica medida na j ésima progênie da i ésima família; m = média geral; g_{ij} = efeito genético aditivo do QTL $\sim N(0, \sigma_g^2)$; a_{ij} = efeito poligênico $\sim N(0, \sigma_a^2)$; e_{ij} = efeito de ambiente $\sim N(0, \sigma_e^2)$.

Todos os efeitos aleatórios do modelo seguem distribuição normal, entretanto, se σ_a^2 e σ_e^2 são suficientemente grandes e normalmente distribuídos, não é exigida a distribuição normal do efeito do QTL.

Para o uso desse modelo, pressupõe-se ausência de interferência e de dominância, e que os *crossing-overs* seguem a distribuição de Poisson e ocorrem à distância de λ Morgan e são uniforme e independentemente distribuídos na região analisada.

A variância de y_{ij} , considerando-se equilíbrio de ligação, é $\text{Var}(y_{ij}) = \sigma^2 = \sigma_g^2 + \sigma_a^2 + \sigma_e^2$ e a covariância entre dois meios-irmãos não-endogâmicos da mesma família é $\text{Cov}(y_{ij}, y_{ij'}) = \pi_q \sigma_q^2 + 1/4 \sigma_a^2$, em que π_q é a proporção de alelos IBD do suposto QTL compartilhada pelos irmãos.

O coeficiente associado à variância poligênica é $1/4$, dado que, em média, espera-se que dois meios-irmãos compartilhem, respectivamente, essas proporções de IBD. As proporções de IBD do QTL (π_q) serão diferentes para cada par de meios-irmãos, podendo variar de 0 a $1/2$.

Para a estimação dos componentes de variância, deve-se substituir π_q por seu valor estimado, $\hat{\pi}_q$. A covariância entre os irmãos j e j' dentro da família i é:

$$V_i = \text{Var} \begin{pmatrix} y_{ij} \\ y_{ij'} \end{pmatrix} = \sigma^2 C_i$$

em que $C_i = \begin{pmatrix} 1 & r_i \\ r_i & 1 \end{pmatrix}$ e $r_i = \hat{\pi}_q h_g^2 + 1/4 h_a^2$.

Com k irmãos em cada família, C_i é uma matriz $k \times k$.

Definindo-se $h_g^2 = \sigma_g^2 / \sigma^2$ como a herdabilidade do suposto QTL, $h_a^2 = \sigma_a^2 / \sigma^2$ como a herdabilidade do componente poligênico, e $h_t^2 = (\sigma_g^2 + \sigma_a^2) / \sigma^2$ como a herdabilidade total e admitindo-se distribuição normal multivariada dos dados, a função densidade conjunta das observações dentro das famílias é:

$$f(y_i) = \frac{1}{(2\pi)^{k/2} |C_i|^{1/2}} * \exp \left\{ -\frac{1}{2\sigma^2} (y_i - v\mu)' C_i^{-1} (y_i - v\mu) \right\}$$

em que $\mathbf{y}_i = [y_{i1}, y_{i2}, y_{i3}, \dots, y_{ik}]'$ é um vetor $\mathbf{k} \times \mathbf{1}$ de valores fenotípicos observados para \mathbf{k} irmãos dentro da família \mathbf{i} , e $\mathbf{v} =$ vetor $\mathbf{k} \times \mathbf{1}$ com todos os elementos iguais a 1.

O log da verossimilhança para \mathbf{n} famílias independentes é:

$$L = \sum_{i=1}^n \ln[f(y_i)]$$

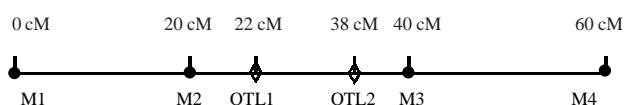
A função de verossimilhança relaciona a posição do QTL flanqueado pelos dois marcadores por meio de \mathbf{r}_i . Os parâmetros desconhecidos a serem estimados são μ , σ^2 , h_g^2 , h_a^2 e θ_{1q} . Para maximizar \mathbf{L} no mapeamento por intervalo, considera-se a fração de recombinação entre o primeiro marcador e o suposto QTL (θ_{1q}) como uma constante, e, gradualmente, eleva-se seu valor e, conseqüentemente, diminuindo-se a distância entre QTL e o marcador posicionado do lado direito (θ_{2q}), em todo o intervalo, repetindo-se o procedimento até que todo o genoma tenha sido coberto. A estimativa de ML da posição do QTL é determinada pelo valor de θ_{1q} que maximiza \mathbf{L} e a hipótese nula $h_g^2 = 0$ pode ser testada por meio do teste de razão de verossimilhança (LR).

Simulação e análises

A técnica de simulação de Monte Carlo foi usada para gerar informações genotípicas e fenotípicas. Para o mapeamento dos QTLs foi considerado um segmento cromossômico de 60 cM com quatro marcadores, igualmente distribuídos no cromossomo a intervalos de 20 cM. Todos os marcadores possuíam seis alelos com a mesma frequência. Dois QTLs, com quatro alelos codominantes com a mesma frequência e efeitos aditivos, foram simulados em três situações:

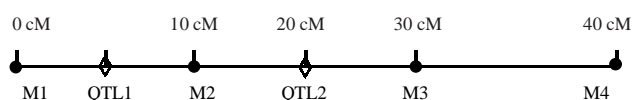
Situação A - Dois QTLs no mesmo intervalo

Situation A - Two QTLs at the same interval



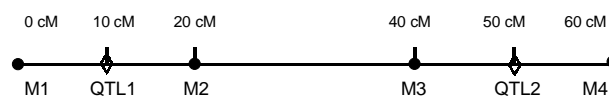
Situação B - Dois QTLs em intervalos adjacentes

Situation B - Two QTLs at adjacent intervals



Situação C - Dois QTLs em intervalos não-adjacentes

Situation C - Two QTLs at no adjacent intervals



Os touros e vacas foram gerados alocando-se aleatoriamente os alelos em cada loco, admitindo-se equilíbrio de Hardy-Weinberg e também que as fases de ligação na geração parental eram desconhecidas. As progênies foram geradas sob suposição de ausência de interferência; deste modo, um evento de recombinação em um intervalo não afeta a possibilidade de ocorrência de outra recombinação no intervalo adjacente. As frequências de recombinação para cada loco foram calculadas usando-se a função de Haldane.

Os dados fenotípicos foram simulados, com distribuição normal, média zero e variância igual a um, sendo gerados a partir do seguinte modelo:

$$y_{ij} = \mu + q_{ij} + \frac{1}{2}(s_i + d_{ij}) + \phi_{ij} + e_{ij}$$

em que y_{ij} = valor fenotípico do $j^{\text{ésimo}}$ indivíduo na $i^{\text{ésima}}$ família de meios-irmãos; μ = constante inerente à cada observação; q_{ij} = efeito dos genótipos dos QTLs do $j^{\text{ésimo}}$ indivíduo; s_i = contribuição do pai para o valor poligênico da progênie; d_{ij} = contribuição da mãe para o valor poligênico da progênie; ϕ_{ij} = efeito da amostragem mendeliana; e_{ij} = erro aleatório.

Para maior simplicidade, não foram considerados efeitos fixos na simulação. Os números de famílias (10 e 50), bem como os tamanhos dessas famílias (10 e 40), foram escolhidos de modo a contemplar diferentes situações possíveis de serem encontradas em populações comerciais de gado de leite. Para maior simplicidade, admitiu-se que os touros não eram aparentados e que as vacas eram tomadas aleatoriamente da população e não possuíam parentesco entre si ou com os touros.

De modo a testar o comportamento do modelo aleatório na detecção de QTLs em gado de leite, foi escolhido o valor igual a 0,25 para a herdabilidade da característica e três diferentes proporções de variância explicada pelos QTLs (20, 40 e 100%), sendo a variância genética total obtida por meio da soma dos componentes referentes aos QTLs e ao efeito poligênico. O número de alelos de cada QTL foi igual a quatro e dos marcadores, seis, sendo suas frequências iguais.

O procedimento de ML no mapeamento por intervalo foi aplicado aos dados simulados, onde o cromossomo de 60 cM foi percorrido a passos de 2 cM da esquerda para a direita. Os parâmetros desconhecidos h_g^2 , h_a^2 e σ^2 foram estimados simultaneamente. A função de verossimilhança foi maximizada em relação a esses parâmetros usando o algoritmo simplex. A posição com o maior LR foi aceita como a mais provável localização do QTL. Para cada combinação de parâmetros (tamanhos e números das famílias, proporção de variâncias dos QTLs etc), a simulação e as análises foram repetidas cem vezes. As acurácias das estimações das posições dos QTLs foram determinadas usando o intervalo de confiança de 95%, estimado a partir da variância entre repetições e calculado como quatro vezes o erro-padrão empírico.

A distribuição empírica do teste LR foi gerada da mesma maneira para cada uma das combinações dos parâmetros sob a hipótese nula de que não havia nenhum QTL no segmento cromossômico. O nível de significância de 0,95 foi escolhido para todas as análises. O valor empírico de limiar foi definido como o 95^o percentil da distribuição empírica do teste LR sob H_0 . O poder de detecção foi definido como o percentual das repetições em que a hipótese nula foi rejeitada ao nível de significância de 5%. As distribuições dos valores máximos obtidos sob H_0 para as três situações estudadas estão ilustradas nas Figuras 1 e 2.

Resultados e Discussão

As estimativas para as localizações dos QTLs, relativas às médias de 100 repetições, com o correspondente intervalo de confiança para as diferentes situações, proporções de variâncias genéticas devidas aos QTLs e tamanhos da amostra são apresentadas na Tabela 1.

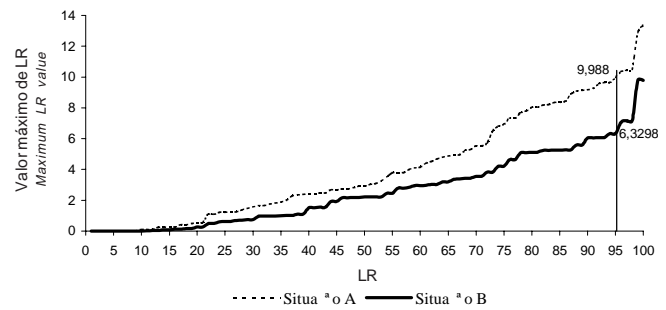


Figura 1 - Valores máximos do teste de razão de verossimilhança (LR) sob H_0 para a Situação A (dois QTL no mesmo intervalo) e Situação B (dois QTLs em intervalos adjacentes).

Figure 1 - Maximum likelihood ratio values (LR) under H_0 for situation A (two QTLs at the same interval) and B (two QTLs at adjacent intervals).

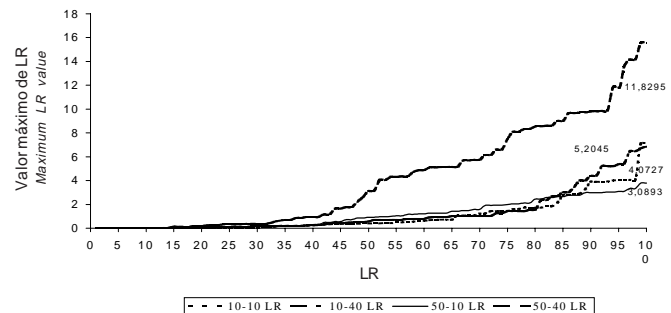


Figura 2 - Valores máximos do teste de razão de verossimilhança sob H_0 da situação C (dois QTLs em intervalos não-adjacentes) para dez famílias com dez irmãos (10-10 LR), para dez famílias com 40 irmãos (10-40 LR), para 50 famílias com dez irmãos (50-10 LR) e para 50 famílias com 40 irmãos (50-40 LR).

Figure 2 - Maximum likelihood ratio values (LR) under H_0 for situation C (two QTLs at non adjacent intervals) using ten families with ten siblings (10-10 LR), ten families with 40 siblings (10-40 LR), 50 families with ten siblings (50-10 LR) and 50 families with 40 siblings (50-40 LR).

Para a Situação A, em que os dois QTLs estão presentes no mesmo intervalo, pode-se verificar que as estimativas das posições foram semelhantes para as duas proporções de variâncias estudadas (40 e 100%). O mapeamento por intervalo, usando uma função de verossimilhança que admitia a presença de somente um QTL no intervalo, não indicou corretamente os picos de LR nos intervalos considerados, localizando um QTL entre os dois simulados, no meio do cromossomo (Figura 3) e também outros dois, um à esquerda do primeiro QTL simulado e outro à direita do segundo QTL. Provavelmente, em virtude de os QTLs estarem a 16 cM de distância, ou seja, muito próximos e sem nenhum marcador entre eles, o método não conseguiu separar adequadamente seus efeitos, sugerindo a existência de três QTLs, formando a “imagem fantasma” (Haley & Knott, 1992). Resultado semelhante foi obtido por Da et al. (2000), os quais verificaram que a utilização do mapeamento por intervalo pode identificar apenas um QTL em determinado intervalo que contenha dois QTLs.

Houve diminuição dos intervalos de confiança, com o aumento da variância de 40 para 100%. É interessante verificar também que, dentre todas as situações e combinações de parâmetros estudadas, a Situação A foi a que obteve os menores erros-padrão e, conseqüentemente, os menores intervalos, provavelmente pela menor distância entre os QTLs.

Na Situação B, em que os QTLs estavam posicionados em intervalos adjacentes, o mapeamento

por intervalo usando modelos aleatórios foi eficiente para localizar corretamente os QTLs somente quando a variância devida aos QTLs explicava 100% da variação genética. Quando a variação foi menor (40%), houve tendência de se indicar, erroneamente, a presença de um QTL entre os dois simulados, próximo à posição 20 cM e outro na posição 60 cM. Uma possível explicação para esse fato, segundo Martinez et al. (1999), pode ser a baixa herdabilidade da característica, aliada à reduzida variação genética explicada pelos QTLs e ao número de famílias usado, tornando as estimativas de posição viesadas.

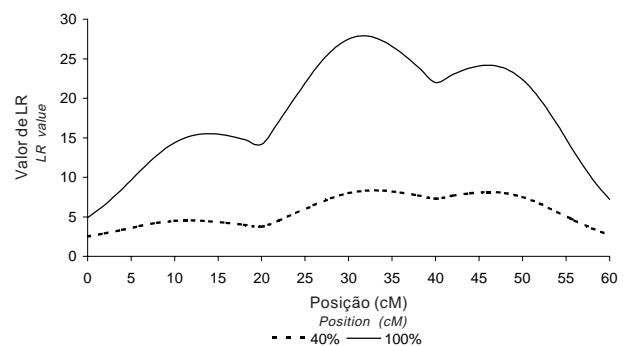


Figura 3 - Comparação dos valores de LR para duas diferentes proporções de variâncias genéticas devidas aos QTLs (40 e 100%), na Situação A.

Figure 3 - Comparison of LR values for two distinct proportions of genetic variances due to QTLs (40 and 100%), in Situation A.

Tabela 1 - Estimativa da posição segundo o tamanho da amostra e a proporção da variância genética devida aos QTLs
Table 1 - Position estimate for sample size and proportion of genetic variation due to QTLs

| Sit ¹ Sit ¹ | NF ² NF ² | TF ³ FS ³ | Proporção da variância genética devida aos QTLs (%) Proportion of genetic variation due to QTLs (%) | | | | | | | | | | | |
|--------------------------------------|------------------------------------|------------------------------------|--|------------------------------------|--------------|----------|--------------|----------|--------------|----------|--------------|----------|--------------|----------|
| | | | 100 | | | | 40 | | | | 20 | | | |
| | | | QTL1 QTL1 | IC ⁴ CI ⁴ | QTL2 QTL2 | IC CI | QTL1 QTL1 | IC CI | QTL2 QTL2 | IC CI | QTL1 QTL1 | IC CI | QTL2 QTL2 | IC CI |
| A | 50 | 40 | 30,7 | 1,2 | 30,7 | 1,2 | 31,3 | 2,5 | 31,3 | 2,5 | - | - | - | - |
| B | 50 | 40 | 9,8 | 1,4 | 29,3 | 2,1 | 20,8 | 5,3 | 59,9 | 6,7 | - | - | - | - |
| | 10 | 10 | 20,5 | 2,8 | 50,6 | 3,1 | 20,6 | 4,9 | 58,7 | 5,1 | 21,6 | 13,6 | 59,0 | 13,2 |
| C | 10 | 40 | 10,5 | 2,6 | 49,2 | 2,7 | 6,7 | 4,8 | 49,3 | 4,9 | 1,6 | 13,0 | 59,4 | 13,1 |
| | 50 | 10 | 21,6 | 2,4 | 55,5 | 2,7 | 19,9 | 4,4 | 59,8 | 4,1 | 17,6 | 10,2 | 60,0 | 11,0 |
| | 50 | 40 | 10,2 | 2,4 | 50,4 | 2,3 | 11,0 | 4,2 | 51,5 | 4,2 | 9,0 | 10,0 | 51,6 | 10,1 |

¹ Sit = situações A (dois QTLs presentes no mesmo intervalo, nas posições 22 e 38 cM), B (dois QTLs em intervalos adjacentes, nas posições 10 e 30 cM) e C (dois QTLs em intervalos não-adjacentes, nas posições 10 e 50 cM); ² NF = número de famílias; ³ TF = tamanho das famílias; ⁴ IC = 4 x erro-padrão empírico.

¹ Sit = situation A (two QTLs at same interval, positions 22 and 38 cM), B (two QTLs at adjacent intervals, positions 10 and 30 cM) and C (two QTLs at no adjacent intervals, positions 10 and 50 cM); ² NF = number of families; ³ FS = family size; ⁴ CI = 4 x empiric standard error.

Verificou-se, ainda, comportamento semelhante ao da Situação A, em relação à redução do intervalo de confiança com o aumento da variância genética.

Quando os QTLs foram simulados em intervalos não-adjacentes, como na Situação C (Tabela 1), as estimativas das posições foram menos viesadas à medida que o número de informações dentro de famílias e a variação genética explicada pelos QTLs aumentavam, culminando na localização quase exata de ambos QTLs, quando havia 50 famílias com 40 indivíduos cada e 100% da variação devida aos QTLs. Observa-se também que, com menor número de informações entre e dentro de famílias (10 famílias com 10 progênies) e reduzidas variações genéticas (20 e 40%), localizaram os QTLs sempre à direita dos simulados, indicando a presença do primeiro QTL na posição 20 cM e do segundo na posição 60 cM (Figura 5). Quando a variação aumentou para 100%, houve tendência de se posicionar os QTLs mais para o centro do cromossomo.

Quando o número de informações dentro de família aumentou de 10 para 40, ocorreu a diminuição do viés na localização dos QTLs, à medida que se aumentou a proporção da variação genética. Entretanto, houve o aparecimento de uma imagem fantasma, sugerindo a existência de um terceiro QTL na posição 30 cM, exatamente entre os dois QTLs simulados (Figura 6). Provavelmente, a utilização de cofatores (Jansen, 1994; Zeng, 1994) pode contornar o problema, absorvendo os efeitos dos

QTLs adjacentes no intervalo investigado. Apesar de ser um método bastante usado na identificação de QTL, o mapeamento por intervalo apresenta alguns problemas, particularmente na distinção de efeitos de múltiplos QTLs ligados. Segundo Knott & Haley (1992), quando dois ou mais QTLs estão localizados em um cromossomo, o mapeamento por intervalo pode produzir estimativas viesadas, localizando erroneamente o QTL. De modo a aumentar a confiabilidade e a acurácia do mapeamento, os efeitos de múltiplos QTLs possivelmente ligados em um cromossomo deveriam ser adequadamente separados para serem testados e estimados (Zeng, 1994). Outras alternativas podem ser usadas, como o mapeamento fino ou a utilização do modelo de múltiplos QTLs (Xu & Atchley, 1995); entretanto, neste último caso, o ganho na acurácia e a perda no poder de detecção têm que ser avaliados.

Em famílias com apenas dez progênies, não houve mudanças expressivas na localização dos QTLs ao serem consideradas dez ou 50 famílias (Figura 7). Todavia, na população em que havia 50 famílias com 40 progênies, o mapeamento por intervalo usando modelos aleatórios tornou-se mais eficiente com o aumento da variação genética explicada pelos QTLs. Da mesma forma que no caso de dez famílias com 40 progênies, observou-se a existência de uma imagem fantasma (Figura 8), localizando um QTL inexistente entre os dois simulados.

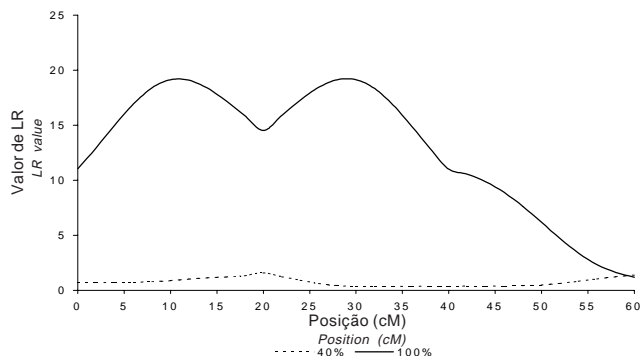


Figura 4 - Comparação dos valores de LR para duas diferentes proporções de variâncias genéticas devidas aos QTLs (40 e 100%), na Situação B.

Figure 4 - Comparison of LR values for two distinct proportions of genetic variances due to QTLs (40 and 100%), in Situation B.

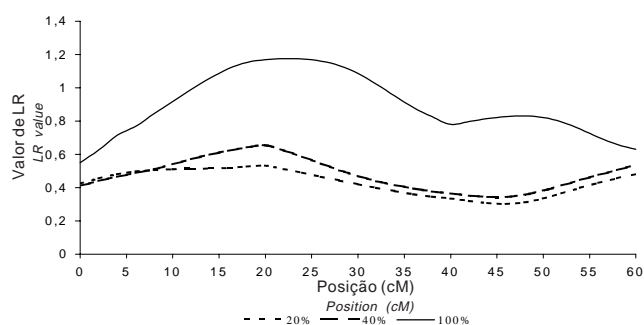


Figura 5 - Comparação dos valores de LR para três diferentes proporções de variâncias genéticas devidas aos QTLs (20, 40 e 100%), na Situação C, para número de família = 10 e tamanho de família = 10.

Figure 5 - Comparison of LR values for three distinct proportions of genetic variances due to QTL (20, 40 and 100%), in Situation C, for a family number = 10 and family size = 10.

Na Situação C, pode ser observado que reduções dos intervalos de confiança ocorreram com os aumentos da variância explicada pelos QTLs e do tamanho da amostra.

As estimativas dos componentes de variância, apresentadas como as herdabilidades do QTL (h_q^2), poligênica (h_a^2) e total (h_t^2) são mostradas na Tabela 2. Considerando as estimativas de h_t^2 , que correspondem às somas de h_q^2 e h_a^2 , observa-se que, de modo geral, estas são próximas ao valor verdadeiro do parâmetro (0,25). As estimativas mais viesadas foram obtidas ao serem analisados os registros de somente dez famílias. Nas análises destes dados, para pequenas variações devidas aos QTLs (20 e 40%), houve subestimação de h_t^2 , independentemente do tamanho de família (10 ou 40), e quando a variação era igual a 100%, ocorreu superestimação de h_t^2 nas famílias com apenas dez indivíduos. Quando o número de famílias aumentava para 50, as estimativas foram muito próximas ao valor paramétrico verdadeiro, mesmo quando a variação devida aos QTLs era de apenas 20%.

Estes resultados são consistentes com os obtidos por Martinez et al. (1998), os quais revelaram que estimativas com maior acurácia são obtidas quando se aumenta o número de animais genotipados.

Embora as estimativas de h_t^2 tenham sido próximas ao valor verdadeiro, não foram obtidas separações adequadas desses valores em seus componentes devidos ao QTL (h_q^2) e poligênico (h_a^2). De acordo com Martinez et al. (1999), a separação desses

componentes pode ser eficientemente obtida somente quando a variação devida aos QTLs explica aproximadamente 10-15% da variância genética. Segundo esses autores, a variância de QTLs de pequeno efeito tende a ser superestimada, enquanto a de QTLs de grandes efeitos é sempre subestimada, com aumento do viés a cada elevação da variação devida aos QTLs. Entretanto, pode ser observado também que a subestimação da variância dos QTLs é sempre acompanhada de superestimação do efeito poligênico, conservando a soma de h_q^2 e h_a^2 próxima do valor verdadeiro, o que pode indicar eficiente separação das variâncias genética e residual (Martinez et al., 1999). Gessler & Xu (1996) afirmaram ainda que o confundimento entre h_q^2 e h_a^2 é considerado uma fragilidade geral do modelo de pares de irmãos.

O poder de detecção empírico, definido como o percentual de repetições nas quais o valor máximo de LR ultrapassou o valor de limiar empírico obtido pela simulação de dados sob H_0 , é apresentado na Tabela 3. O poder de detecção foi altamente dependente da proporção da variância genética, do número de famílias e do tamanho destas famílias. Nas situações e combinações de parâmetros cujo número de famílias ou de indivíduos era dez, o poder de detecção foi igual a zero, independentemente da variância devida aos QTLs. Valores satisfatórios para o poder de detecção (acima de 70%) só foram obtidos ao serem analisados registros de 50 famílias com 40 indivíduos cada, onde os QTLs eram responsáveis por toda a variação genética.

Tabela 2 - Estimativas de herdabilidade devido aos QTLs (h_q^2), poligênica (h_a^2) e total (h_t^2), de acordo com o tamanho da amostra e a proporção da variância genética devida aos QTLs

Table 2 - Estimates of heritability due to QTL (h_q^2), polygenic (h_a^2) and total (h_t^2), according to sample size and the proportion of genetic variation due to QTLs

| Situação <i>Situation</i> | NF <i>NF</i> | TF <i>FS</i> | Proporção da variância genética devida aos QTLs (%) <i>Proportion of genetic variation due to QTLs (%)</i> | | | | | | | | |
|------------------------------|-----------------|-----------------|---|---------|---------|---------|---------|---------|---------|---------|---------|
| | | | 100 | | | 40 | | | 20 | | |
| | | | h_q^2 | h_a^2 | h_t^2 | h_q^2 | h_a^2 | h_t^2 | h_q^2 | h_a^2 | h_t^2 |
| A | 50 | 40 | 0,12 | 0,19 | 0,31 | 0,06 | 0,20 | 0,26 | - | - | - |
| B | 50 | 40 | 0,10 | 0,16 | 0,26 | 0,04 | 0,20 | 0,24 | - | - | - |
| | 10 | 10 | 0,17 | 0,12 | 0,29 | 0,11 | 0,11 | 0,22 | 0,10 | 0,09 | 0,19 |
| C | 10 | 40 | 0,09 | 0,17 | 0,26 | 0,04 | 0,16 | 0,20 | 0,03 | 0,16 | 0,19 |
| | 50 | 10 | 0,12 | 0,13 | 0,25 | 0,07 | 0,07 | 0,18 | 0,05 | 0,19 | 0,24 |
| | 50 | 40 | 0,09 | 0,17 | 0,26 | 0,04 | 0,04 | 0,20 | 0,02 | 0,22 | 0,24 |

¹NF = número de famílias; ²TF = tamanho das famílias.

¹NF = number of families; ²FS = family size.

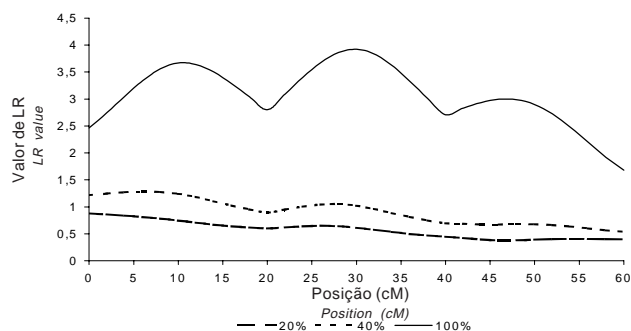


Figura 6 - Comparação dos valores de LR para três diferentes proporções de variâncias genéticas devidas aos QTLs (20, 40 e 100%), na Situação C, para número de família = 10 e tamanho de família = 40.

Figure 6 - Comparison of LR values for three distinct proportions of genetic variances due to QTLs (20, 40 and 100%), in Situation C, for a family number = 10 and family size = 40.

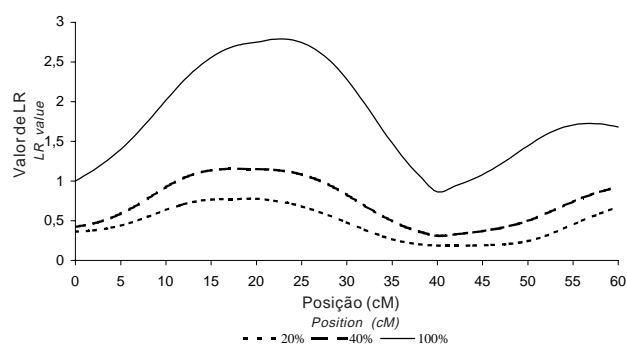


Figura 7 - Comparação dos valores de LR para três diferentes proporções de variâncias genéticas devidas aos QTLs (20, 40 e 100%), na Situação C, para número de família = 50 e tamanho de família = 10.

Figure 7 - Comparison of LR values for three distinct proportions of genetic variances due to QTLs (20, 40 and 100%), in Situation C, for a family number = 50 and family size = 10.

Tabela 3 - Poder de detecção para cada situação, de acordo com o tamanho da amostra e a proporção da variância genética devido aos QTLs

Table 3 - Power of detection for each situation according to sample size and proportion of genetic variance due to QTLs

| Situação Situation | NF ¹ NF | TF ² FS | Limiar Threshold | Poder de detecção Power of detection | | |
|-----------------------|-----------------------|-----------------------|---------------------|--|----|----|
| | | | | Proporção da variância genética devido aos QTLs (%) Proportion of genetic variation due to QTLs (%) | | |
| | | | | 100 | 40 | 20 |
| A | 50 | 40 | 9,88 | 83 | 0 | - |
| B | 50 | 40 | 6,33 | 81 | 0 | - |
| | 10 | 10 | 4,07 | 0 | 0 | 0 |
| C | 10 | 40 | 5,41 | 0 | 0 | 0 |
| | 50 | 10 | 3,10 | 0 | 0 | 0 |
| | 50 | 40 | 11,83 | 72 | 0 | 0 |

¹NF = número de famílias; ²TF = tamanho das famílias.

¹NF = number of families; ²FS = family size.

Martinez et al. (1999) afirmaram que o poder para a detecção de QTLs, em geral, é altamente dependente da herdabilidade da característica e da proporção da variância genética explicada por ele, enquanto outros estudos, como os realizados por Blackwelder & Elston (1982) e Amos et al. (1989), mostraram que o tamanho da amostra e o número de alelos do marcador são também fatores importantes que contribuem para o aumento do poder de detecção.

Conclusões

O mapeamento por intervalo baseado em modelos aleatórios foi eficiente na estimação da localização de QTLs, em grandes famílias de meios-irmãos, quando esses eram responsáveis por 100% da variação genética e não estavam presentes no mesmo intervalo. Entretanto, o método parece não localizar com eficiência os QTLs de pequenos efeitos, principalmente se

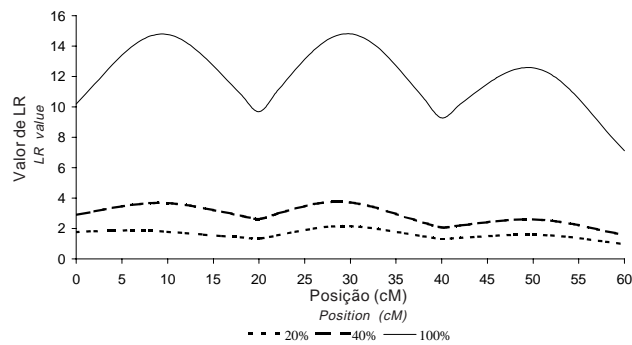


Figura 8 - Comparação dos valores de LR para três diferentes proporções de variâncias genéticas devidas aos QTLs (20, 40 e 100%), na Situação C, para número de família = 50 e tamanho de família = 40.

Figure 8 - Comparison of LR values for three distinct proportions of genetic variances due to QTLs (20, 40 and 100%), in Situation C, for a family number = 50 and family size = 40.

estes estiverem localizados próximos às extremidades do cromossomo.

Por ser um método simples e robusto, o modelo aleatório pode ser recomendado para a varredura rápida do genoma, aplicando-se, em seguida, análises mais refinadas nos intervalos em que for indicada a existência de um suposto QTL.

O modelo aleatório não consegue separar eficientemente a variância poligênica e a dos QTLs, devendo ser usados métodos mais sofisticados para estimar essa última.

Literatura Citada

- AMOS, C.I.; ELSTON, R.C. Robust methods for detection of genetic linkage for quantitative data from pedigrees. **Genetic Epidemiology**, v.6, p.349-360, 1989.
- AMOS, C.I.; ELSTON, R.C.; WILSON, A.F. et al. A more powerful robust sib-pair test of linkage for quantitative traits. **Genetic Epidemiology**, v.6, p.435-449, 1989.
- BLACKWELDER, W.C.; ELSTON, R.C. Power and robustness of sib-pair linkage tests and extension to larger sibships. **Communication Statistics Theory and Methods**, v.11, p.449-484, 1982.

- DA, Y.; VAN RADEN, P.M.; SCHOOK, L.B. Detection and parameter estimation for quantitative trait loci using regression models and multiple markers. **Genetics Selection Evolution**, v.32, p.357-381, 2000.
- FULKER, D.W.; CARDON, L.R. A sib-pair approach to interval mapping of quantitative trait loci. **American Journal of Human Genetics**, v.47, p.957-967, 1994.
- GESSLER, D.D.G.; XU, S. Using the expectation or the distribution of the identity by descent for mapping quantitative trait loci under the random model. **American Journal of Human Genetics**, v.59, p.1382-1390, 1996.
- GOLDGAR, D.E. Multipoint analysis of human quantitative genetic variation. **American Journal of Human Genetics**, v.47, p.957-967, 1990.
- HALEY, C.S.; KNOTT, S.A. A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. **Heredity**, v.69, p.315-324, 1992.
- HASEMAN, J.K.; ELSTON, R.C. The investigation of linkage between a quantitative trait and a marker locus. **Behavior Genetics**, v.2, p.3-19, 1972.
- JANSEN, R.C. Controlling the Type I and Type II errors in mapping quantitative trait loci. **Genetics**, v.138, p.871-881, 1994.
- KADARMIDEEN, H.N.; DEKKERS, J.C.M. Regression on markers with uncertain allele transmission for QTL mapping in half-sib designs. **Genetics Selection Evolution**, v.31, p.437-455, 1999.
- KNOTT, S.A.; HALEY, C.S. Maximum likelihood mapping of quantitative trait loci using full-sibs families. **Genetics**, v.132, p.1211-1222, 1992.
- MARTINEZ, M.L.; VUKASINOVIC, N.; FREEMAN, A.E. et al. Mapping QTL in outbred populations using selected samples. **Genetics Selection Evolution**, v.30, p.453-468, 1998.
- MARTINEZ, M.L.; VUKASINOVIC, N.; FREEMAN, A.E. Random model approach for QTL mapping in half-sib families. **Genetics Selection Evolution**, v.31, p.319-340, 1999.
- MARTINEZ, M.L.; VUKASINOVIC, N. Algoritmo para cálculo da proporção de genes idênticos por descendência, para mapear QTL em famílias de meio-irmãos. **Revista Brasileira de Zootecnia**, v.29, p.443-451, 2000.
- XU, S.; ATCHLEY, W.R. A random model approach to interval mapping of quantitative trait loci. **Genetics**, v.141, p.1189-1197, 1995.
- ZENG, Z.-B. Precision mapping of quantitative trait loci. **Genetics**, v.136, p.1457-1468, 1994.

Recebido em: 01/09/03

Aceito em: 13/10/04