

UTILIZAÇÃO DE MODELOS ALEATÓRIOS NA ESTIMAÇÃO DA LOCALIZAÇÃO DE QTL EM FAMÍLIAS DE MEIOS-IRMÃOS

AUTORES

MARCOS VINICIUS GUALBERTO BARBOSA DA SILVA^{1,2} MÁRIO LUIZ MARTINEZ^{1,2} ROBLEDO DE ALMEIDA TORRES^{2,3} PAULO SÁVIO LOPES^{2,3} RICARDO FREDERICO EUCLYDES², MARCO ANTÔNIO MACHADO¹, WAGNER ARBEX⁴

¹ Pesquisador da Embrapa Gado de Leite– Rua Eugênio do Nascimento, 610 – 36038-330 – Juiz de Fora - MG

² Bolsista do CNPq

³ Professor Adjunto do DZO/UFV

⁴ Técnico de Nível Superior da Embrapa Gado de Leite

RESUMO

O mapeamento por intervalo baseado em modelo aleatório foi aplicado para estudar sua robustez e propriedades no mapeamento de dois QTL 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. As proporções de genes idênticos por descendência (IBD) dos QTL foram estimadas a partir das proporções IBD de dois marcadores flanqueadores. As estimativas dos parâmetros dos QTL (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 e o tamanho das famílias, a proporção da variância genética devido aos QTL e à posição dos QTL, localizados no mesmo intervalo, em intervalos adjacentes e não-adjacentes. Os fatores que mais influenciaram as estimativas dos parâmetros foram a proporção da variância devido aos QTL, o número e o tamanho das famílias. As estimativas mais viesadas dos componentes dos QTL e poligênico foram obtidas ao serem analisados os registros de somente dez famílias. Com número suficiente de famílias e de indivíduos nas famílias e altas proporções de variância genética, devido aos QTL, o modelo aleatório pode detectar QTL com alto poder, apresentando estimativas das posições com boa acurácia, se não existem dois QTL no mesmo intervalo.

PALAVRAS-CHAVE

mapeamento por intervalo, marcador genético, método dos pares de irmãos

TITLE

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 a random model, QTL location and variance components were estimated using maximum likelihood techniques. The estimation of parameters was based on the sib-pair approach. The proportions of genes identical-by-descent (IBD) at the two QTL were estimated from the IBD at two flanking marker loci. Estimates for QTL parameters (locations and variance components) and power of detection, in a trait with $h^2 = 0.25$, were obtained using simulated data, and varying the number of families, number of half-sibs by families, proportion of QTL variance and QTL positions. These QTL were positioned at the same interval, at adjacent intervals and at no adjacent intervals. The most important factors influencing the estimates of QTL parameters and power were the proportion of variance due to QTL, the number and size of the half-sibs families. The most biased variance components estimations were obtained when the sample size was ten. Given a sufficient number of families and high proportions of genetic variance due to QTL, the random model approach can detect a QTL with high power and provides accurate estimates of the QTL position if there are no two QTL in the same interval.

KEYWORDS

genetic marker, interval mapping , sib-pair method

INTRODUÇÃO

A identificação e o mapeamento de genes que influenciam as características quantitativas, chamados de quantitative trait loci ou simplesmente QTL, têm recebido muita atenção nos últimos anos (KADARMIDEEN e DEKKERS, 1999), principalmente com o desenvolvimento de mapas de ligação com grande número de marcadores moleculares. O objetivo, nesse caso, é acelerar o progresso genético com o uso dos QTL identificados.

Para a identificação desses locos, geralmente, são utilizados dados provenientes de cruzamentos entre linhagens endogâmicas (ou raças), que incluem, por exemplo, o retrocruzamento e o F2, ou de populações chamadas segregantes ou exogâmicas (XU e ATCHLEY, 1995). Os dados provenientes das populações segregantes podem ser obtidos de modo mais fácil e barato; entretanto, apresentam como desvantagens o fato dos 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.

De acordo com AMOS e ELSTON (1989), métodos robustos, que utilizam modelos aleatórios, baseados no compartilhamento de diferentes proporções de genes idênticos por descendência (IBD) entre dois parentes, vêm sendo desenvolvidos e não requerem a especificação de modelos genéticos para a detecção da ligação. Desta forma, objetivou-se com este estudo testar a robustez do modelo aleatório no mapeamento de dois QTL 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 QTL, em característica com herdabilidade igual a 0,25.

MATERIAL E MÉTODOS

A técnica de simulação de Monte Carlo foi usada para gerar informações genotípicas e fenotípicas. Para o mapeamento de dois QTL 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 QTL, com quatro alelos codominantes com a mesma frequência e efeitos aditivos, foram simulados em três situações: A) dois QTL no mesmo intervalo; B) dois QTL em intervalos adjacentes; C) dois QTL em intervalos nãoadjacentes.

Os dados fenotípicos foram simulados, com distribuição normal, média zero e variância igual a um, a partir de um modelo que incluiu os efeitos do genótipo do QTL, da contribuição do pai e da mãe para o valor poligênico da progênie, da amostragem mendeliana e do erro. As progênies foram geradas sob suposição de ausência de interferência e as frequências de recombinação foram calculadas usando-se a função de Haldane. Nas análises foi usado o modelo proposto por GOLDGAR (1990), sob a pressuposição de ausência de interferência e de dominância, e que os crossing-overs seguiam a distribuição de Poisson e eram uniforme e independentemente distribuídos na região analisada. A variância fenotípica era igual a soma dos componentes do efeito genético do QTL (σ_q^2), do poligênico (σ_a^2) e do erro (σ_e^2). A covariância entre dois meios-irmãos não-endogâmicos da mesma família foi considerada como $Cov(y_{ij}, y_{ij}') = \pi_q \sigma_q^2 + \frac{1}{4} \sigma_a^2$, onde π_q é a proporção de alelos IBD dos supostos QTL (MARTINEZ e VUKASINOVIC, 2000).

Não foram considerados efeitos fixos na simulação. Os números de famílias (10 e 50), e os tamanhos dessas famílias (10 e 40) foram escolhidos de modo a contemplar diferentes situações possíveis em populações comerciais de gado de leite. Ainda, foi escolhido o valor igual a 0,25 para a herdabilidade da característica e três diferentes proporções de variância explicada pelo QTL (20, 40 e 100%). 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. A função de verossimilhança foi maximizada usando o algoritmo

simplex. A posição com o maior LR foi aceita como a mais provável localização do QTL. Para cada combinação de parâmetros, a simulação e as análises foram repetidas cem vezes. As acurácias das estimações das posições dos QTL foram determinadas a partir da variância entre repetições e calculada como quatro vezes o erro-padrão empírico. O valor empírico de limiar foi definido como o 95º percentil da distribuição empírica do teste LR sob a hipótese nula de que não havia nenhum QTL no segmento cromossômico.

RESULTADOS E DISCUSSÃO

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

Na Situação A, em que os dois QTL estão presentes no mesmo intervalo, verifica-se que o mapeamento por intervalo não indicou corretamente os picos de LR, localizando um QTL entre os dois simulados, e também outros dois, um à esquerda do primeiro QTL simulado e outro à direita do segundo QTL. Provavelmente, por estarem muito próximos e sem nenhum marcador entre eles, o método não conseguiu separar adequadamente os efeitos dos QTL, sugerindo a existência de três QTL, formando a "imagem fantasma", como apresentado na Figura 1 (HALEY e KNOTT, 1992). Com o aumento da variância de 40 para 100%, ocorreu diminuição dos intervalos de confiança.

Na Situação B (QTL posicionados em intervalos adjacentes), o método foi eficiente para localizar corretamente os QTL quando a variância devido ao QTL explicava 100% da variação genética. Quando a variação foi menor (40%), houve a tendência de indicar, erroneamente, a presença de um QTL entre os dois simulados. Uma possível explicação para esse fato, segundo MARTINEZ et al. (1999), pode ser a baixa herdabilidade, aliada à reduzida variação genética explicada pelos QTL e ao número de famílias usado, tomando as estimativas de posição viesadas. Verificou-se comportamento semelhante ao da Situação A, em relação à redução do intervalo de confiança com o aumento da variância genética.

Quando os QTL foram simulados em intervalos não-adjacentes (Situação C), as estimativas das posições foram menos viesadas à medida que o número de informações dentro de famílias e a variação genética explicada pelos QTL aumentavam, culminando na localização quase exata de ambos QTL quando havia 50 famílias com 40 indivíduos cada uma e 100% da variação devido ao QTL. Houve, todavia, o aparecimento de uma imagem fantasma, sugerindo a existência de um terceiro QTL na posição 30 cM, exatamente entre os dois QTL simulados. Provavelmente, a utilização de cofatores (JANSEN, 1994; ZENG, 1994) pudesse contornar o problema, absorvendo os efeitos dos QTL adjacentes, no intervalo sendo investigado. Com menor número de informações entre e dentro de famílias e reduzidas variações genéticas, os QTL foram localizados sempre à direita da posição simulada.

Apesar de ser o método padrão na identificação de QTL, o mapeamento por intervalo apresenta alguns problemas, particularmente na distinção de efeitos de múltiplos QTL ligados. Segundo KNOTT e HALEY (1992), quando dois ou mais QTL 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 QTL possivelmente ligados 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 QTL (XU e ATCHLEY, 1995); entretanto, o ganho na acurácia e a perda no poder tem que ser avaliados.

CONCLUSÕES

O mapeamento por intervalo baseado em modelos aleatórios pode ser aplicado em populações de gado de leite para a identificação de QTL que sejam responsáveis por grande proporção da variação genética, podendo ser detectados com alto poder e localizados com exatidão, principalmente em grandes famílias de meio-irmãos e se não estiverem presentes no mesmo intervalo.

REFERÊNCIAS BIBLIOGRÁFICAS

1. 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.
2. GOLDFAR, D.E. . Multipoint analysis of human quantitative genetic variation. *American Journal of Human Genetic*, v. 47, p.957:967, 1990.
3. 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.
4. JANSEN, R.C. . Controlling the Type I and Type II errors in mapping quantitative trait loci. *Genetics*, v. 138, p. 871-881, 1994.
5. KADARMIDEEN, H.N., DEKKERS, J.C.M. . Regression on markers with uncertain allele transmission for QTL mapping in half-sib designs. *Genetic Selection Evolution*, v. 31, p.437-455, 1999.
6. KNOTT, S.A., HALEY, C.S. . Maximum likelihood mapping of quantitative trait loci using full-sibs families. *Genetics*, v. 132, p. 1211-1222, 1992.
7. MARTINEZ, M.L., VUKASINOVIC, N., FREEMAN, A. E. . Random model approach for QTL mapping in half-sib families. *Genetic Selection. Evolution*, v. 31, p.319-340, 1999.
8. 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.
9. XU, S., ATCHLEY, W.R. . A random model approach to interval mapping of quantitative trait loci. *Genetics*, v. 141, p.1189-1197, 1995.
10. ZENG, Z.-B.. Precision mapping of quantitative trait loci. *Genetics*, v. 136, p. 1457-1468, 1994.

TABELA 1. Estimativa da posição segundo o tamanho da amostra e a proporção da variância genética devido ao QTL.

Sit ¹	NF ²	TF ³	Proporção da variância genética devido ao QTL (%)											
			100				40				20			
			QTL1	IC ⁴	QTL2	IC	QTL1	IC	QTL2	IC	QTL1	IC	QTL2	IC
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 QTL presentes no mesmo intervalo, nas posições 22 e 38 cM), B (QTL em intervalos adjacentes, nas posições 10 e 30 cM) e C (QTL 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.

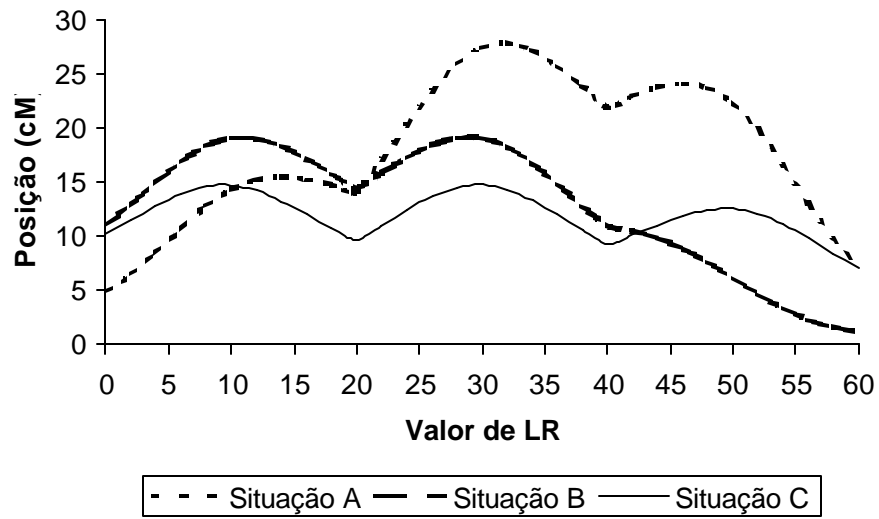


FIGURA 1. Comparação dos valores de LR para as Situações A, B e C para proporção da variância genética devido ao QTL = 100%, número de famílias = 50 e tamanho de família = 40.