

# ESTIMATIVAS DE COMPONENTES DE VARIÂNCIA DE QTL E PODER DE DETECÇÃO EM FAMÍLIAS DE IRMÃOS COMPLETOS USANDO MODELOS ALEATÓRIOS

## AUTORES

MARCOS VINICIUS GUALBERTO BARBOSA DA SILVA<sup>1,2</sup> MÁRIO LUIZ MARTINEZ<sup>1,2</sup> ROBLEDO DE ALMEIDA TORRES<sup>2,3</sup> PAULO SÁVIO LOPES<sup>2,3</sup> RICARDO FREDERICO EUCLYDES<sup>2</sup>, MARCO ANTÔNIO MACHADO<sup>1</sup>, WAGNER ARBEX<sup>4</sup>

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

<sup>2</sup> Bolsista do CNPq

<sup>3</sup> Professor Adjunto do DZO/UFV

<sup>4</sup> Técnico de Nível Superior da Embrapa Gado de Leite

## RESUMO

O poder estatístico e os componentes de variância relativos a um delineamento de filhas para detecção da ligação entre marcadores e QTL foram estimados a partir de simulações de diferentes estruturas de populações. As populações foram simuladas variando-se o número e o tamanho de famílias, as variâncias devido ao QTL e as posições de dois QTL em um cromossomo de 60 cM, com seis marcadores espaçados a cada 20 cM. Para tanto, utilizou-se o método da máxima verossimilhança no mapeamento por intervalo, sob modelo aleatório. Os resultados indicaram que o modelo aleatório pode ser aplicado para a varredura rápida do genoma.

## PALAVRAS-CHAVE

marcador genético, máxima verossimilhança, modelo aleatório, QTL

## TITLE

ESTIMATING QTL VARIANCE COMPONENTS AND POWER IN FULL-SIBS FAMILIES UNDER THE RANDOM MODEL

## ABSTRACT

Statistical power and variance components of an experimental daughter design for detecting linkage among markers and QTL were estimated for some situations by simulation. The experimental populations were generated from different structures varying the number and size of the families and the QTL variance in a 60 cM long chromosomal segment covered by six markers, equally distributed along the chromosome at a 20 cM distance from each other. Maximum likelihood techniques were applied in the interval mapping under random model. The results indicated that the random model can be applied for genome screening.

## KEYWORDS

genetic marker, maximum likelihood, random model, QTL

## INTRODUÇÃO

O modelo clássico de herança, baseado na aditividade dos genes, provou ser robusto em médio prazo, pois bons resultados têm sido obtidos com sua utilização. Contudo, em situações onde os efeitos maiores sobre o fenótipo são devidos a um ou a poucos pares de genes, este modelo não tem mostrado a mesma eficiência. Tais genes de efeito maior, segregando em uma população, são chamados de quantitative trait loci, ou simplesmente QTL (BOICHARD, 1998)

Atualmente, são conhecidos três métodos de análises de QTL em populações segregantes: os baseados em modelos fixos, em modelos mistos ou em modelos aleatórios, podendo ser categorizados com base no modelo

genético-aditivo adotado e na informação utilizada. Para alguns dos modelos, procedimentos como regressão, máxima verossimilhança (MARTINEZ et al., 1998) e técnica de Monte Carlo em Cadeia de Markov (BINK e VAN ARENDONK, 2000) podem ser utilizados para estimação de médias, variâncias e da taxa de recombinação entre os marcadores genéticos e o QTL.

No melhoramento animal, o emprego de modelos aleatórios no mapeamento de QTL parece ser vantajoso, em razão das estruturas de famílias encontradas, como em gado de leite, onde existem grandes famílias de meias-irmãs e a coleta de dados fenotípicos é rotineira. Seu uso é baseado na similaridade fenotípica entre indivíduos geneticamente relacionados, sendo a (co)variância entre dois indivíduos da mesma família expressa pelos componentes do efeito poligênico e do QTL. O primeiro depende apenas do parentesco genético aditivo entre os animais, enquanto o segundo depende da proporção de genes idênticos por descendência (IBD) compartilhada por eles. Assim, objetivou-se estimar os componentes de variância e o poder de detecção em famílias de irmãos completos por meio de modelos aleatórios.

## MATERIAL E MÉTODOS

GOLDGAR (1990) descreveu um modelo aleatório que, na sua forma geral, pode ser escrito como  $y_{ij} = \mu + g_{ij} + a_{ij} + e_{ij}$  onde  $y_{ij}$  = valor observado da característica medido na  $j$ ésima progênie da  $i$ ésima família;  $\mu$  = média da população;  $g_{ij}$  = efeito genético aditivo do QTL  $\sim N(0, \sigma_g^2)$ ;  $a_{ij}$  = efeito aditivo de todos os genes, com exceção do QTL testado  $\sim N(0, \sigma_a^2)$ ;  $e_{ij}$  = efeito de ambiente  $\sim N(0, \sigma_e^2)$ . Para o uso desse modelo, pressupõe-se ausência de interferência e de dominância, os crossing-overs seguem a distribuição de Poisson, ocorrem a uma distância de  $\lambda$  Morgans, e são uniforme e independentemente distribuídos na região analisada. A variância de  $y_{ij}$ , assumindo-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 irmãos completos não-endogâmicos, pertencentes à mesma família é  $\text{Cov}(y_{ij}, y_{ij'}) = \pi_q \sigma_g^2 + \frac{1}{2} \sigma_a^2$ , onde  $\pi_q$  é a proporção de alelos IBD do suposto QTL compartilhada pelos irmãos.

Definindo-se  $h_g^2 = \frac{\sigma_g^2}{\sigma^2}$  como a herdabilidade do suposto QTL,  $h_a^2 = \frac{\sigma_a^2}{\sigma^2}$  como a herdabilidade do componente poligênico, e  $h_t^2 = \frac{(\sigma_g^2 + \sigma_a^2)}{\sigma^2}$  como a herdabilidade total e admitindo-se distribuição normal multivariada dos dados,

tem-se que o log da verossimilhança para  $n$  famílias de irmãos completos independentes é  $L = \sum_{i=1}^n \ln[f(y_i)]$ . Para maximizar  $L$  no mapeamento por intervalo, considera-se a fração de recombinação entre o primeiro marcador e o suposto QTL ( $\theta_{1q}$ ) como uma constante, e, gradualmente, vai-se aumentando seu valor e, conseqüentemente, diminuindo-se a distância entre QTL e o marcador posicionado do lado direito ( $\theta_{2q}$ ), em todo o intervalo, repetindo-se o procedimento até que todo o genoma tenha sido coberto. A estimativa ML da posição do QTL é determinada pelo valor de  $\theta_{1q}$  que maximiza  $L$  e a hipótese nula  $h_g^2 = 0$  pode ser testada por meio do teste de razão de verossimilhança (LR). 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 estimativas dos componentes de variância foram obtidas a partir de dados simulados, como descrito por SILVA (2002).

## RESULTADOS E DISCUSSÃO

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 1. As estimativas de  $h_t^2$ , que correspondem às somas de  $h_q^2$  e  $h_a^2$ , foram, de modo geral, viesadas em relação ao valor verdadeiro do parâmetro (0,25). Para pequenas variações devido ao QTL, houve superestimação de  $h_t^2$ , independente do tamanho de família e quando a variação era igual a 100%, ocorreu subestimação de  $h_t^2$  em praticamente todas as famílias, independente do número de indivíduos, exceto quando o número de famílias aumentava para 50 e o número de indivíduos era igual a 10.

Também não foram obtidas separações adequadas dos componentes devido ao QTL ( $h_q^2$ ) e

poligênico ( $h_a^2$ ). MARTINEZ et al. (1999) afirmaram que a variância de QTL de pequeno efeito tendem a ser superestimadas, enquanto a de QTL de grandes efeitos é sempre subestimada, com aumento do viés a cada elevação da variação devido ao QTL. Entretanto, pode ser observado também que a subestimação da variância do QTL é sempre acompanhada da superestimação do efeito poligênico; todavia, para pequenas amostras e para QTL posicionados no mesmo intervalo ou em intervalos adjacentes, os valores obtidos pelas somas de  $h_q^2$  e  $h_a^2$  são menores quando comparados ao valor verdadeiro, indicando que, nesses casos, não ocorrem separações eficientes das variâncias genética e residual. Como afirmado por GESSLER e XU (1996), 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 2. Independentemente da variância devido ao QTL, o poder foi igual a zero para as situações e combinações de parâmetros cujo número de famílias ou de indivíduos era cinco, mesmo resultado obtido para variâncias do QTL iguais a 20 e 40%. Nas Situações A e B, onde havia 50 famílias com dez indivíduos, o poder, apesar de superior a zero, foi baixo, o que pode indicar que as posições dos QTL no cromossomo estão relacionadas ao poder de detecção. Ainda, ao se distribuir a variância genética para dois QTL talvez ocorra alguma perda no poder de detecção, quando comparado às situações onde somente um suposto QTL seja identificado, como nos trabalhos de VAN DER BEEK (1996) e MARTINEZ et al. (1999). Neste caso, talvez número maior de repetições seja necessário para estimar o poder com maior precisão e assegurar que os resultados não reflitam flutuações estatísticas. Outro ponto a ser destacado é que se admitiu que os marcadores possuíam frequências iguais na população. Provavelmente, o poder estatístico seria reduzido, caso as frequências fossem diferentes em razão da influência do QTL sobre a média da característica, pois um dos genótipos poderá ter maior frequência e, desse modo, tornar assimétrica a distribuição normal da característica.

O poder para a detecção foi altamente dependente do tamanho da amostra, das posições e da proporção da variância genética explicada pelos QTL, pois somente quando havia 50 famílias de 10 indivíduos cada uma, QTL posicionados em intervalos não-adjacentes e toda a variação explicada por eles, o poder foi próximo ao considerado como satisfatório (70%). A dependência da proporção da variância foi também observada por MARTINEZ et al. (1999), enquanto outros estudos, como os realizados por BLACKWELDER e ELSTON (1982) e AMOS et al. (1989), mostraram que o tamanho da amostra e o número de alelos do marcador são também fatores importantes e que contribuem para o aumento do poder de detecção. Elevações no poder de detecção com o aumento do número de famílias de irmãos completos e de progênies por família também foram obtidas por VAN DER BEEK (1996).

## CONCLUSÕES

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 as variâncias poligênicas e dos QTL, devendo ser usados métodos mais sofisticados para estimar essa última.

## 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. BINK, M., VAN ARENDONK, J.A.M. Detection of quantitative trait loci in outbred population with incomplete

- marker data. *Genetics*, v. 151, p. 409-420, 2000.
3. BLACKWELDER, W.C., ELSTON, R.C. Power and robustness of sib-pair linkage tests and extension to larger sibships. *Commun. Statist. Theor. Methods*, v. 11, p. 449-484, 1982.
  4. BOICHARD, D. QTL detection with genetic markers in dairy cattle: an overview. In: 49<sup>th</sup> Annual Meeting of the EAAP, Varsóvia, Polônia, p. 1-6, 1998.
  5. GESSLER, 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.
  6. 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.
  7. MARTINEZ, M.L., VUKASINOVIC, N., FREEMAN, A.E. Estimating QTL location and QTL variance in half-sibs families under the random model with missing parental genotypes. *Animal Breeding and Genetics*, v. 115, p. 165-180, 1998.
  8. SILVA, M.V.G.B. Utilização de modelos aleatórios no mapeamento de quantitative trait loci em famílias de irmãos completos e de meios-irmãos. Viçosa, MG:UFV, 2002. 125 p. Tese (Doutorado em Genética e Melhoramento) – Universidade Federal de Viçosa, 2002.
  9. VAN DER BEEK, S. The use of genetic markers in poultry breeding. Wageningen: The Netherlands, 1996, 175p. Tese (Doutorado em Melhoramento Animal) – Wageningen University, The Netherlands, 1996.

Tabela 1. Estimativas de herdabilidade devido ao QTL ( $h_q^2$ ), poligênica ( $h_a^2$ ) e total ( $h_t^2$ ), de acordo com o tamanho da amostra e a proporção da variância genética devido ao QTL.

Situação	NF	TF	Proporção da variância genética devido ao QTL (%)								
			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	10	0,12	0,09	0,21	0,08	0,30	0,38	-	-	-
B	50	10	0,11	0,10	0,21	0,04	0,34	0,38	-	-	-
	10	5	0,08	0,15	0,23	0,10	0,30	0,40	0,10	0,32	0,42
C	10	10	0,08	0,13	0,21	0,06	0,21	0,27	0,06	0,24	0,30
	50	5	0,09	0,09	0,18	0,11	0,23	0,33	0,11	0,26	0,37
	50	10	0,12	0,13	0,25	0,05	0,35	0,40	0,03	0,43	0,46

<sup>1</sup>NF = número de famílias; <sup>2</sup>TF = tamanho das famílias

Tabela 2. Poder de detecção para cada situação, de acordo com o tamanho da amostra e a proporção da variância genética devido ao QTL.

Situação	NF <sup>1</sup>	TF <sup>2</sup>	Limiar	Poder de detecção		
				Proporção da variância genética devido ao QTL (%)		
A	50	10	6,08	13	0	-
B	50	10	5,38	11	0	-
	10	5	3,86	0	0	0
C	10	10	4,01	0	0	0
	50	5	6,38	0	0	0
	50	10	2,58	68	0	0

<sup>1</sup>NF = número de famílias; <sup>2</sup>TF = tamanho das famílias