



## ESTIMATIVAS DE COMPONENTES DE VARIÂNCIA DE QTL E PODER DE DETECÇÃO EM FAMÍLIAS DE MEIOS-IRMÃOS

MARCOS VINICIUS GUALBERTO BARBOSA DA SILVA<sup>1</sup>, GUSTAVO HENRIQUE DE SOUZA<sup>2</sup>, MÁRIO LUIZ MARTINEZ<sup>1</sup>, ROBLEDO DE ALMEIDA TORRES<sup>3</sup>, PAULO SÁVIO LOPES<sup>3</sup>, MARCO ANTÔNIO MACHADO<sup>1</sup>, ANA LUISA SOUZA AZEVEDO<sup>4</sup>

<sup>1</sup> Estudante de Doutorado DZO-UFV

<sup>2</sup> Pesquisador da Embrapa Gado de Leite, Rua Eugênio do Nascimento, 610, CEP 36038-330, Juiz de Fora-MG

<sup>3</sup> Professor da Universidade Federal de Viçosa – DZO, CEP 36570-000, Viçosa-MG

<sup>4</sup> Bolsista da FAPEMIG

**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, o número de alelos e as variâncias do QTL, em um cromossomo de 100 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 poder de detecção foi bastante reduzido quando há poucos indivíduos em pequenas famílias, mesmo para QTL de grande efeito.

**PALAVRAS-CHAVE:** genes idênticos por descendência, modelo aleatório, QTL

### ESTIMATING QTL VARIANCE COMPONENTS AND POWER IN HALF-SIBS FAMILIES

**ABSTRACT** – Statistical power and variance components of daughter design experiment for the detection of linkage among markers and QTL were estimated by simulation in different families structures. The experimental populations were generated from different structures through variation of the number and size of the families and the number of alleles and the QTL variance in a 100 cM long chromosomal segment covered by six markers equally distributed along the chromosome with 20 cM distance from each other. Maximum likelihood techniques were applied in interval mapping under a random model. Results indicated that the power of detection was low when the data had few and small families.

**KEYWORDS:** identical by descent genes, QTL, random model

### INTRODUÇÃO

Vários métodos têm sido implementados para a localização e mapeamento de QTLs. As diferenças entre eles, basicamente, estão relacionadas aos requerimentos computacionais, melhor utilização das informações disponíveis (propriedades distribucionais, *pedigree* etc.) e acurácia no mapeamento. Dentre esses, métodos robustos, baseados em modelos aleatórios, vêm sendo desenvolvidos (Fulker e Cardon, 1994), os quais 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 IBD, maior é 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 devido ao efeito do QTL. O componente poligênico é composto pela soma dos efeitos de todos os locos que influenciam a característica, com exceção do suposto QTL. Apresenta como principais vantagens a aplicação em *pedigrees* complexos, robustez para número maior de alelos do QTL e o desconhecimento da fase de ligação dos marcadores. Assim, objetivou-se estimar os componentes de variância e o poder de detecção de um QTL em famílias de meios-irmãos por meio de modelos aleatórios.

### MATERIAL E MÉTODOS

Um modelo aleatório pode ser escrito como  $y_{ij} = \mu + g_{ij} + a_{ij} + e_{1j}$  onde  $y_{ij}$  = valor observado da característica medido na  $j^{\text{ésima}}$  progênie da  $i^{\text{ésima}}$  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 meios-irmãos 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 meios-irmãos 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%. Assim, foram simuladas populações considerando um cromossomo de 100 cM, com a presença de um QTL a 50 cM e de seis marcadores espaçados a cada 20 cM, sendo a herdabilidade da característica igual a 0,25 e o número de famílias igual a 10. Para cada combinação dos parâmetros foram realizadas 100 repetições por tratamento, variando-se o número de filhas (25 e 50), o número de alelos por QTL (2, 4, 6 e 10) e a variação genética devida ao QTL (10, 50 e 100%).

### RESULTADOS E DISCUSSÃO

As estimativas dos componentes de variância, apresentadas como herdabilidades do QTL ( $h_q^2$ ), poligênica ( $h_a^2$ ) e total ( $h_t^2$ ), estão na Tabela 1. Ressalte-se que as estimativas de  $h_t^2$  correspondem às somas de  $h_q^2$  e  $h_a^2$ .

Para o tamanho de família igual a 50 indivíduos, todas as estimativas de  $h_t^2$  foram subestimadas e não houve diferença entre elas, apesar da existência de variação no número de alelos do QTL. Aumentando-se a proporção da variância genética devida ao QTL houve aproximação do valor estimado em relação ao valor verdadeiro (0,25). O valor mais próximo do real ocorreu com o número de alelos do QTL igual a 2 e toda a variância genética devida ao QTL.

Para tamanho de família igual a 25, as estimativas de  $h_t^2$  foram subestimadas quando a variância genética devida ao QTL foi igual a 10%, e superestimada com 100% da variância genética devida ao QTL. Quando 50% da variância genética era devida ao QTL, as estimativas foram próximas ao valor real do parâmetro para 2 e 4 alelos e exatamente iguais para 6 e 10 alelos. O mesmo comportamento das estimativas de  $h_t^2$  foi obtido por Silva (2003) quando o número de famílias era de apenas 10 indivíduos.

Silva (2003) obteve estimativas de  $h_t^2$  muito próximas ao valor paramétrico verdadeiro, mesmo quando a variação devida ao QTL era de apenas 20%, para famílias compostas por 50 indivíduos. De acordo com esse autor, os resultados obtidos foram consistentes com os relatados por Martinez et al. (1998), os quais revelaram que estimativas com maior acurácia podem ser obtidas quando o número de animais genotipados é aumentado.

Neste estudo, para 100 famílias de 50 indivíduos cada uma e 100% da variância genética devida ao QTL, o valor da estimativa de  $h_t^2$  foi igual ao valor paramétrico verdadeiro (0,25), semelhante ao relatado por Martinez et al. (1998).

Mesmo quando as estimativas de  $h_t^2$  foram iguais ou próximas ao valor verdadeiro, não foram obtidas separações adequadas desses valores em seus componentes devido 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 devido ao QTL explica aproximadamente 10-15% da variância genética. Segundo esses autores, as variâncias de QTL de pequeno efeito tendem a ser superestimadas, enquanto a de QTL de grandes efeitos é sempre subestimada, com o aumento do viés a cada elevação da variância devido ao QTL. Entretanto pode ser observado, também, que a subestimação da variância do QTL 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 genéticas e residual (Martinez et al., 1999). Gessler e Xu (1996) afirmaram também que o confundimento entre  $h_q^2$  e  $h_a^2$  é considerado fragilidade do modelo aleatório.

TABELA 1. Efeito do número de alelos do QTL, tamanho da família e proporção da variância genética devido ao QTL sobre as estimativas de herdabilidade devido ao QTL ( $h_q^2$ ), poligênica ( $h_a^2$ ) e total ( $h_t^2$ )

NF	TF	NA	Proporção da variância genética devido ao QTL (%)								
			10			50			100		
			$h_q^2$	$h_a^2$	$h_t^2$	$h_q^2$	$h_a^2$	$h_t^2$	$h_q^2$	$h_a^2$	$h_t^2$
10	25	2	0,05	0,16	0,22	0,08	0,16	0,23	0,12	0,17	0,29
		4	0,05	0,17	0,22	0,07	0,16	0,24	0,11	0,17	0,28
		6	0,05	0,17	0,22	0,08	0,17	0,25	0,12	0,17	0,28
		10	0,06	0,17	0,22	0,08	0,17	0,25	0,12	0,17	0,29
10	50	2	0,02	0,14	0,16	0,03	0,13	0,17	0,05	0,19	0,23
		4	0,02	0,14	0,16	0,03	0,14	0,17	0,04	0,17	0,21
		6	0,02	0,14	0,16	0,03	0,14	0,17	0,04	0,16	0,20
		10	0,02	0,14	0,16	0,03	0,13	0,16	0,04	0,16	0,20
100	50	4	-	-	-	-	-	-	0,08	0,17	0,25

NF = número de famílias TF = Tamanho da família; NA = número de alelos do QTL.

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  $H_0$ , foi igual a zero, independentemente da variância devido ao QTL, tamanho de família, número de alelos do QTL. Os mesmos valores foram obtidos por Silva (2003), quando o número de famílias foi igual a 10, o qual afirma que o poder de detecção foi altamente dependente da proporção da variância genética devido ao QTL, número de famílias e tamanho de família e que valores satisfatórios para o poder de detecção (acima de 70%) só foram obtidos com 50 famílias de 40 indivíduos cada, onde o QTL era responsável por toda a variância genética.

Martinez et al. (1999) afirmaram que o poder de detecção é dependente da herdabilidade da característica e da proporção da variância genética explicada por ele. Outros estudos, como o realizado por 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 poder de detecção.

### CONCLUSÕES

Os resultados indicaram que o poder de detecção foi bastante reduzido quando há poucos indivíduos em pequenas famílias, mesmo para QTL de grande efeito.

Ainda, o número de alelos do QTL parece não influenciar na estimação da localização ou na estimação dos componentes de variância.



**REFERÊNCIAS BIBLIOGRÁFICAS**

- AMOS, C. I.; ELSTON, R. C. Robust methods for detection of genetic linkage for quantitative data from pedigrees. **Genet. Epidemiol.**, v. 6, p. 349-360, 1989.
- FULKER, D. W; CARDON, L. R. A sib-pair approach to interval mapping of quantitative trait loci. **Am. J. Hum Genet.**, 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. **Am. J. Hum. Genet.**, v. 59, p. 1382-1390, 1996.
- 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. **Anim. Breed. Genet.**, v. 115, p. 165-180, 1998.
- MARTINEZ, M. L.; VUKASINOVIC, N.; FREEMAN, A. E. Random model approach for QTL mapping in half-sib families. **Genet. Sel. Evol.**, v. 31, p. 319-340, 1999.
- SILVA, M. V. G. B. **Utilização de modelos aleatórios no mapeamento de QTL em famílias de meio-irmãos e irmãos completos**. 2003. 99 f. Tese (Doutorado em Genética e Melhoramento) – Universidade Federal de Viçosa, Viçosa, MG.