

GUSTAVO HENRIQUE DE SOUZA

**DETECÇÃO DE QTLS EM FAMÍLIAS DE MEIOS-IRMÃOS POR MEIO
DE GENES IDÊNTICOS POR DESCENDÊNCIA, UTILIZANDO
SIMULAÇÃO**

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento, para obtenção do título de “Magister Scientiae”.

VIÇOSA
MINAS GERAIS - BRASIL
2003

GUSTAVO HENRIQUE DE SOUZA

**DETECÇÃO DE QTLS EM FAMÍLIAS DE MEIOS-IRMÃOS POR MEIO
DE GENES IDÊNTICOS POR DESCENDÊNCIA, UTILIZANDO
SIMULAÇÃO**

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento, para obtenção do título de “Magister Scientiae”.

APROVADA: 26 de fevereiro de 2003.

Prof. Ricardo Frederico Euclides

Dr. Mário Luiz Martinez
(Conselheiro)

Prof^a. Carmen Silva Pereira

Prof^a. Simone Guimarães

Prof. Robledo de Almeida Torres
(Orientador)

A Deus
A minha esposa Milian,
Aos meus pais José Onício e Maria do Carmo,
As minhas irmãs Débora, Daniele e Gisele,
Aos demais familiares

AGRADECIMENTO

A Deus, pela existência e consciência da necessidade de sempre seguir em frente.

À Universidade Federal de Viçosa e ao Departamento de Genética e Melhoramento, pela oportunidade de realização do curso.

À Conselho Nacional de Pesquisa (CNPq), pela concessão da bolsa de estudos.

Ao professor Robledo de Almeida Torres, pela orientação, pela amizade, pela confiança e pelo apoio a todas as atividades do curso.

Ao professor Paulo Sávio Lopes e ao Dr. Mário Luiz Martinez, pelo aconselhamento, pela amizade e pela atenção dispensada sempre que solicitada.

Ao pesquisador da EMBRAPA - Gado de Leite, Dr. Marcos Vinicius Gualberto Barbosa da Silva, pelo Sistema GeneSibs, sugestões e críticas apresentadas.

Às professoras Carmem Silva Pereira e Simone Guimarães, pelas sugestões e pelas críticas apresentadas.

A minha esposa Milian, pelo amor e pelo carinho sempre presentes em nossas vidas.

Aos familiares, pelo apoio e pela amizade.

A amiga Rachel, pelo apoio, dias de estudo na biblioteca e pelo convívio saudável durante todo o mestrado.

Aos amigos do “Melhoramento” pelo apoio, pelo agradável convívio e pela amizade.

Aos demais professores, funcionários, alunos do Departamento de Zootecnia e de genética e melhoramento e a todos que, direta ou indiretamente, contribuíram para a realização deste trabalho.

BIOGRAFIA

Gustavo Henrique de Souza, filho de José Onício de Souza e Maria do Carmo Oliveira de Souza, natural de Viçosa, Minas Gerais, nasceu em 29 de novembro de 1977.

Em fevereiro de 1996, iniciou o curso de Zootecnia pela Universidade Federal de Viçosa e no período de agosto de 1997 a julho de 2000 foi bolsista de Iniciação Científica pelo Pibic/CNPq sob a orientação dos professores Ricardo Frederico Euclides e Robledo de Almeida Torres ambos do departamento de Zootecnia.

Em março de 2001 graduou-se em Zootecnia pela Universidade Federal de Viçosa e, em abril de 2001, iniciou o curso de Mestrado em Genética e Melhoramento pela Universidade Federal de Viçosa submetendo-se ao exame final de defesa de tese em fevereiro de 2003.

INDICE

	Página
RESUMO.....	vi
ABSTRACT.....	viii
1 - INTRODUÇÃO.....	1
2 - REVISÃO DE LITERATURA.....	4
2.1 - Delineamentos experimentais em grandes populações segregantes ..	4
2.1.1 - Delineamento de Filhas.....	6
2.1.2 - Única família de meias-irmãs.....	7
2.1.3 - Várias famílias de meias-irmãs.....	9
2.2 - Métodos fundamentados em modelos aleatórios.....	12
3 - MATERIAL E MÉTODOS.....	18
3.1 – SIM_HS 1Q.....	19
3.2 – IBD_HS_1Q.....	21
3.3 – EST_HS_1Q.....	24
4 - RESULTADOS E DISCUSSÃO.....	29
5 – CONCLUSÕES.....	40
6 - REFERÊNCIAS BIBLIOGRÁFICAS.....	41

RESUMO

SOUZA, Gustavo Henrique de, M.S., Universidade Federal de Viçosa, fevereiro de 2003. **Detecção de QTLs em famílias de meios-irmãos por meio de genes idênticos por descendência, utilizando simulação.** Orientador: Robledo de Almeida Torres. Conselheiros: Paulo Sávio Lopes e Mário Luiz Martinez.

Para muitas espécies de animais domésticos, árvores frutíferas perenes e especialmente seres humanos, a formação de linhagens endogâmicas é impraticável, sendo uma alternativa para o mapeamento de QTL a execução de análises de marcadores moleculares entre famílias existentes dentro de uma população segregante. O procedimento de mapeamento por intervalo baseado em modelo aleatório foi aplicado para estudar a robustez e propriedades no mapeamento de QTL, em populações simuladas, constituídas de famílias de meios-irmãos, variando o número de famílias, o tamanho da família, o número de alelos do QTL e a percentagem da variância genética explicada pelo QTL. As famílias de meios-irmãos foram simuladas considerando um segmento cromossômico de 100 cM, com seis marcadores igualmente distribuídos no cromossomo, a intervalos de 20 cM, com seis alelos de mesma frequência. O número de alelos do suposto QTL foi de 2, 4, 6 ou 10, sempre com as mesmas frequências. O número de famílias foi de 10 ou 100 e os tamanhos de famílias foram de 25 ou 50 com 10, 50 ou 100 % da proporção da variância genética devida ao QTL. Sob o modelo aleatório, a localização e os componentes de variância foram estimados, usando o método da máxima verossimilhança. As estimativas dos parâmetros foram baseadas na metodologia de pares de irmão. A proporção de genes idênticos por descendência (IBD) do QTL foi estimada pelo algoritmo proposto por MARTINEZ e VUKASINOVIC (2000). Na estimação

dos parâmetros do QTL, que são a localização e os componentes de variância observou-se pouca influencia do número de alelos do QTL, sendo a proporção da variância genética devida ao QTL e ao número de indivíduos analisados, o que mais influenciou nas estimativas. Quando havia apenas 10 famílias, as estimativas de localização dos componentes de variância do QTL e da variância poligênica foram sempre viesadas. O mapeamento por intervalo estabelecido em modelos aleatórios em populações de meios-irmãos para a identificação de QTL parece não ser apropriado para investigar a presença de QTL, quando existem apenas 10 famílias com 25 ou 50 filhos. O número de alelos do QTL parece não influenciar na estimação da localização ou na estimação dos componentes de variância. O método indicou corretamente o intervalo do suposto QTL. Por ser simples e robusto, o modelo aleatório pode ser recomendado para a varredura rápida do genoma, aplicando-se, em seguida, análises mais refinadas no intervalo em que for indicada a existência de um suposto QTL. O modelo aleatório não consegue separar eficientemente a variância poligênica e as variâncias dos QTLs, propondo-se o uso de métodos mais sofisticados para estimar as variâncias devidas ao QTL.

ABSTRACT

SOUZA, Gustavo Henrique de, M.S., Universidade Federal de Viçosa, February, 2003. **Detection of QTLs in half-sibs' families through genes identical by descent, using simulation.** Adviser: Robledo de Almeida Torres. Committee Members: Paulo Sávio Lopes and Mario Luiz Martinez.

For many species of domestic animals, perennial fruit trees and especially in human beings, the formation of inbred lines is impracticable, being an alternative for the mapping of QTL the execution of analyses of molecular markers among existent families of a outbred population. The mapping procedure by interval based on random model was applied to study the robustness and properties in the mapping of QTL in simulated populations of half-sibs families, varying the number of families, the family size, number of QTL allele and the percentage of the genetic variance explained by QTL. The half-sibs families were simulated considering a chromosomal segment of 100 cM, with six markers equally distributed in the chromosome at intervals of 20 cM, with six allele of same frequency. The number of allele of the putative QTL was 2, 4, 6 or 10 allele, always with the same frequencies. The number of families was 10 or 100 and the sizes of families were 25 or 50 with 10, 50 or 100% of the proportion of the genetic variance due to QTL. Under the random model the location and the variance components were estimated using the maximum likelihood method. The estimates of the parameters were based on the brother's pairs methodology. The proportion of identical-by- descent (IBD) of QTL was estimated by the algorithm proposed by MARTINEZ and VUKASINOVIC (2000). In the estimate of the QTL parameters, which are location and variance components, it was observed little influences of the number of allele of QTL being the proportion of the genetic variance due to QTL and the number of analyzed individuals what more

influenced in the estimates. When there had been only 10 families, the location estimates of the components of QTL variance and polygenic variance were always biased. The mapping by interval based on random models in half-sibs populations for the identification of QTL is not appropriate to investigate the presence of QTL when there are only 10 families with 25 or 50 children. The number of allele of QTL does not influence in the estimate of the location or in the estimate of the variance components. The method indicated correctly the interval of the putative QTL. Since it is simple and robust, the random model can be recommended for a fast sweeping of the genome, being applied, soon after, more refined analyses in the interval that the existence of a putative QTL is indicated. The random model is not able to separate the polygenic variance and the QTL variance efficiently, being proposed the use of more sophisticated methods to estimate the variances due to QTL.

1 – INTRODUÇÃO

Em programas de melhoramento, as características de importância econômica como peso, produção de leite, resistência a doenças, entre outras, são exemplos de características quantitativas (ou QT - *quantitative traits*). A genética quantitativa estuda a herança destas características baseada na suposição de que elas são determinadas por vários genes, cujos efeitos são aditivos e sofrem influência do ambiente a que estão expostos.

Durante muitos anos, os estudos em genética quantitativa baseavam-se somente na observação do fenótipo para prever os parâmetros genéticos das populações (herdabilidade, variâncias genética aditiva, ambientais, etc.) e os valores genéticos dos indivíduos. Hoje, a ciência volta-se novamente para a questão básica da herança, de uma maneira mais sofisticada, tentando mapear, ou seja, identificar e posicionar os genes que contribuem para a expressão das características quantitativas (QT's). Alguns destes locos, cujos efeitos são maiores sobre as QT's, são chamados de *quantitative trait loci* ou simplesmente QTLs.

O desenvolvimento das técnicas de análise de DNA, de diversos tipos de marcadores moleculares (*Restriction fragment length polymorphisms*-RFLP, mini e microsatélites, etc.) e da reação em cadeia de polimerase (PCR) propiciou o início de projetos em larga escala com a elaboração de mapas de ligação com grande número de marcadores moleculares, espalhados por todo o genoma em humanos e em várias espécies de animais, como suínos, bovinos, ovinos, caprinos, etc..

O desenvolvimento de mapas de ligação com grande número de marcadores tem estimulado a procura por métodos que permitam mapear genes

responsáveis por variação em características quantitativas e verificar as possíveis associações entre QTLs, marcadores e fenótipos.

A detecção de QTL com o auxílio de marcadores requer a existência de desequilíbrio de ligação entre sua seqüência e o locus marcador (BOILARD e BOICHARD, 2002), que é usado para definir a herança conjunta de alelos em locos diferentes, ou seja, se dois locos estão próximos (ligados) em um mesmo cromossomo, há maior probabilidade de eles serem herdados juntos.

O mapeamento de QTL pode ser dividido em três fases: a detecção de fatores genéticos que tenham efeito sobre a característica de interesse econômico e estejam segregando na população; a localização do QTL em relação ao marcador e, por fim, a estimação do efeito do QTL e suas interações.

O mapeamento de QTLs baseia-se em testes de associações entre marcadores moleculares e dados fenotípicos, por meio de várias metodologias estatísticas. De modo a aumentar o poder de detecção e reduzir o viés na estimação da posição e do efeito do QTL, diversos delineamentos têm sido propostos, envolvendo o cruzamento entre linhagens endogâmicas, que diferem no método utilizado para se “criarem” ou identificarem pais heterozigotos para os locos dos marcadores e QTLs e os cruzamentos realizados. Para muitas espécies de animais domésticos, árvores frutíferas perenes e especialmente seres humanos, a formação de linhagens endogâmicas é impraticável. Desta forma, uma alternativa é a execução de análises em famílias existentes dentro de uma população segregante (MARTINEZ, 1998).

Segundo SILVA (2003), atualmente, pode-se optar entre três métodos de análises, em populações segregantes: métodos baseados em modelos fixos, em modelos mistos ou em modelos aleatórios. Estes métodos são categorizados com base no modelo genético-aditivo adotado compreendendo o número de alelos do QTL, o número de efeitos genéticos, etc., e na informação utilizada na análise. Para alguns dos modelos, diferentes procedimentos, tais como regressão (KADARMIDDEN e DEKKERS, 1999), máxima verossimilhança (MARTINEZ et al., 1998) e *Markov Chain Monte Carlo* (HOESCHELE et al. 1997; BINK e VAN ARENDONK, 2000) podem ser utilizados para estimação de médias e

variâncias dos genótipos de QTL, bem como da taxa de recombinação entre os marcadores genéticos e o QTL.

Em situações típicas de melhoramento animal, o emprego de modelos aleatórios parece ser freqüentemente mais vantajoso, em razão das estruturas de famílias encontradas, como, por exemplo, em gado de leite, onde existem grandes famílias de meios-irmãos, em virtude do emprego da inseminação artificial e coleta de dados fenotípicos como rotina (SILVA, 2003).

O objetivo desta pesquisa é avaliar a eficiência da detecção de QTLs em famílias de meios-irmãos, por meio de genes idênticos por descendência, utilizando modelos aleatórios, a partir de informações provenientes de famílias simuladas de meios-irmãos simuladas variando o número de filhos, o número de alelos do QTL e a percentagem da variação genética explicada pelo QTL, obtidos através de um pacote computacional para simulação e detecção de QTLs (GeneSibs).

2 – REVISÃO DE LITERATURA

2.1 - Delineamentos experimentais em grandes populações segregantes

A maioria das análises realizadas para se detectar QTL, segundo MARTINEZ (1998), tem sido baseada em cruzamentos planejados, embora alguns estudos tenham sido realizados com populações existentes de animais, especialmente em gado de leite. O primeiro passo para a obtenção das populações é cruzar duas linhagens que diferem quanto aos marcadores genéticos, de modo a produzir progênies heterozigotas F_1 . Após este passo, as progênies dos seguintes tipos de acasalamento têm sido consideradas para análise:

1. Genotipagem seletiva e testes de progênie;
2. Linhagens endogâmicas recombinantes (RIL);
3. Linhagens haplóides duplicadas (DHL);
4. Linhagens isogênicas próximas(NIL);
5. Linhagens avançadas intercruzadas (AIL);
6. Retrocruzamento (BC) de indivíduos F_1 com uma das linhagens parentais;
7. Indivíduos F_2 , obtidos por cruzamento entre F_1 's ou autofecundação.

Para muitas espécies de animais domésticos, árvores frutíferas perenes e, especialmente seres humanos, a formação de linhagens endogâmicas, que formam a base dos delineamentos descritos anteriormente, é impraticável. Desta forma, uma alternativa é a execução de análises em famílias existentes dentro de uma população (MARTINEZ, 1998). Todavia, como os experimentos para o mapeamento de QTLs envolvem a genotipagem e a mensuração fenotípica de um grande número de indivíduos (SOLLER e GENIZI, 1978), torna-se necessário

lançar mão de delineamentos experimentais, de forma a diminuir o custo e garantir razoável poder de detecção.

De acordo com LYNCH e WALSH (1998), o cruzamento entre as populações exogâmicas (*outbred populations*) pode ser de interesse por outras razões, que não as restrições experimentais. Segundo estes autores, QTLs detectados por cruzamentos entre linhagens endogâmicas podem representar diferenças entre as linhagens utilizadas, permanecendo obscura a relevância desses resultados na segregação de QTLs dentro da população endogâmica. No entanto, é de fundamental importância, observar que os métodos são complementares e não, competidores.

LYNCH e WALSH (1998) afirmaram também, que a utilização da variação dentro da população exogâmica resulta em significativa diminuição do poder de detecção. Com o uso de delineamentos envolvendo linhagens endogâmicas, todos os pais F_1 possuirão genótipos idênticos, incluindo a mesma fase de ligação. Assim, todos os indivíduos serão informativos e o desequilíbrio será máximo, ao passo que, em populações exogâmicas, somente uma fração dos pais (ou avôs) será heterozigota para o QTL e para o marcador, sendo desconhecida a fase de ligação. Neste caso, os indivíduos poderão diferir quanto à fase de ligação, podendo assim, o alelo **M** para o marcador estar ligado ao alelo **Q** do QTL em alguns pais e com o alelo **q** em outros, o que demonstrando a necessidade de se efetuarem as análises considerando-se os pais individualmente. Além disso, na população F_1 , haverá somente dois alelos segregando para as linhagens endogâmicas, enquanto um número indefinido de alelos pode estar segregando na população exogâmica.

Outro ponto importante, que está ligado ao poder de detecção, são os feitos dos QTLs serem expressos como médias nas linhagens endogâmicas e como variâncias genéticas nas populações exogâmicas, sendo essas variâncias estimadas com menor precisão.

O desequilíbrio de ligação entre marcador e QTL pode ser encontrado dentro de famílias, em razão da segregação conjunta (BOVENHUIS *et al.*, 1997). Em gado de leite, os delineamentos mais utilizados são o Delineamento de Filhas

(*Daughter Design*) ou de Netas (*Granddaughter Design*), dentro de famílias de meias-irmãs paternas.

2.1.1 - Delineamento de Filhas

WELLER *et al.* (1990) relataram que o mapeamento de QTLs por meio de marcadores se baseia na esperança de que, se um indivíduo é heterozigoto para o locus do marcador e do QTL, então, a menos que ocorra recombinação, a progênie que recebe um determinado alelo do marcador, receberá também o QTL ligado. Conseqüentemente, quando as progênies são agrupadas de acordo com o alelo do marcador recebido a partir de um pai heterozigoto, a presença de diferentes alelos do QTL ligados tenderá a gerar uma diferença entre as médias dos dois grupos, em relação à característica quantitativa de interesse, que, se for significativa, indicará a presença do QTL próximo ao marcador.

NEIMANN-SORENSEN e ROBERTSON (1961) propuseram o **delineamento de filhas** (*Daughter design*), que é usado principalmente em gado de leite, em que pode ser obtido um grande número de progênies por touro, principalmente pelo uso intensivo de inseminação artificial, sendo a presença de ligação, testada pela estatística **t** aplicada à diferença entre as médias dos dois grupos de marcadores.

Neste delineamento, há a necessidade de se genotipar, além do touro, todas as suas filhas, de modo a classificá-las de acordo com o marcador, bem como mensurar seus registros fenotípicos, podendo tornar o experimento mais caro. Duas situações são mais comuns. Na primeira, são avaliados os registros de meios-irmãos, filhos de um único touro; na segunda, é avaliado o conjunto de dados disponíveis que se referem às famílias de meios-irmãos, filhos de vários touros. A seguir serão abordadas as duas situações.

2.1.2 - Única família de meias-irmãs

Considere que um touro heterozigoto para o marcador e QTL seja acasalado com n vacas (não-aparentadas e sem informação sobre os marcadores), sendo obtida, de cada acasalamento, somente uma filha, que é mensurada para a característica de interesse. Denotando os alelos do marcador, presentes no touro, como M_1 e M_2 , pode ser utilizado o seguinte modelo linear para a verificação da ligação entre o marcador e o QTL:

$$Y_{ij} = \mu + M_i + e_{ij}$$

onde

Y_{ij} = valor da característica para a $j^{\text{ésima}}$ filha com o $i^{\text{ésimo}}$ alelo do marcador;

μ = média da população;

M_i = efeito do $i^{\text{ésimo}}$ genótipo do marcador;

e_{ij} = erro aleatório $\sim N(0, \sigma^2)$.

Como existem somente duas classes de marcadores, a estatística t pode ser usada para determinar a significância do contraste, sendo dada por

$$t_M = \frac{\mu_{M1} - \mu_{M2}}{\sqrt{\text{Var}(\mu_{M1})\text{Var}(\mu_{M2})}} = \frac{\text{contraste}}{\text{erro} - \text{padrão}}, \text{ com } N-2 \text{ graus de liberdade,}$$

onde N é o número de filhas avaliadas.

Um valor significativo para o teste revela a ligação entre o marcador e o QTL, o qual é heterozigoto no touro.

Supondo-se que a frequência de recombinação seja igual a r , as probabilidades condicionais para os alelos do QTL, dado que o alelo do marcador para o touro seja conhecido, são

$$\begin{array}{ll} P(Q_1|M_1) = 1-r & P(Q_2|M_1) = r \\ P(Q_1|M_2) = r & P(Q_2|M_2) = 1-r \end{array}$$

Assumindo-se também, que os mesmos alelos Q_1 e Q_2 do QTL estejam segregando na população de vacas com frequências iguais a p e $1-p$, respectivamente, e que o acasalamento seja realizado ao acaso, podem-se estimados os valores esperados para as classes de marcadores e para o contraste entre eles (Tabela 1).

Tabela 1. Probabilidades de ocorrência dos genótipos no delineamento de filhas e valores esperados da característica quantitativa, segundo a classe do marcador

Alelo paterno do marcador	Alelo paterno do QTL	Probabilidade do alelo paterno do QTL	Alelo materno do QTL	Probabilidade do alelo materno do QTL	Valor da característica na progênie	Valor esperado para o genótipo do marcador
M_1	Q_1	$1-r$	Q_1	p	a	
M_1	Q_1	$1-r$	Q_2	$1-p$	d	
M_1	Q_2	r	Q_1	p	d	$a(p-r)+$
M_1	Q_2	r	Q_2	$1-p$	$-a$	$d(1-r-p+2rp)$
M_2	Q_1	r	Q_1	p	a	
M_2	Q_1	r	Q_2	$1-p$	d	
M_2	Q_2	$1-r$	Q_1	p	d	$a(r+p-1)+$
M_2	Q_2	$1-r$	Q_2	$1-p$	$-a$	$d(p+r-2rp)$
Contraste					$a(1-2r)+d(1-2r)(1-2p)$	

A expressão relativa à média (ou valor esperado) da classe de marcador M_1 , é obtida como:

$$\text{Média de } M_1 = \mu_{M_1} = a(p-r) + d(1-r-p+2rp)$$

Para a obtenção da média de M_2 , o procedimento é similar. Fazendo-se $\mu_{M_1} - \mu_{M_2}$, obtém-se o contraste, que se for significativo, revela a existência de ligação entre marcador e QTL, dentro da família analisada. Observa-se que, se a

freqüência dos dois alelos do QTL for igual a 0,5, o contraste torna-se $\mathbf{D} = \mathbf{a} (1-2r)$. Por outro lado, se a ligação for total ($r=0$), o contraste torna-se $\mathbf{D} = \mathbf{a}+\mathbf{d} (1-2p)$. Ao se definir a freqüência do segundo alelo como $q = 1 - p$, o contraste se torna igual a $\mathbf{D} = \mathbf{a}+\mathbf{d}(q-p)$, que é a fórmula para o efeito de substituição gênica (FALCONER e MACKAY, 1996).

2.1.3 - Várias famílias de meias-irmãs

No intuito de se aumentar o poder de detecção, em casos onde há limitado número de filhas por reprodutor, a análise de conjuntos de dados contendo informações de várias famílias de touros, torna-se interessante. Outra razão para a confecção deste tipo de análise, é que em algumas situações, um touro pode ser heterozigoto para o marcador e homozigoto para o QTL. Conseqüentemente, em uma análise desta família, a diferença entre as médias das classes de marcadores das progênies deste touro seria igual a zero.

Do mesmo modo, em uma população em equilíbrio, onde 50% dos indivíduos são heterozigotos, o valor esperado para o contraste entre as famílias seria igual a zero, impossibilitando a simples combinação de contrastes entre as famílias. Desta forma, ao serem efetuadas análises de dados provenientes de várias famílias de meias-irmãs, é necessário levar em consideração que a ligação entre marcador e QTL pode diferir entre os touros. O modelo apropriado, nesta situação, seria

$$Y_{ijk} = \mu + S_i + M_{ij} + e_{ijk}$$

onde

Y_{ijk} = valor da característica para a $k^{\text{ésima}}$ filha com o $j^{\text{ésimo}}$ alelo do marcador do $i^{\text{ésimo}}$ touro;

μ = média da população;

S_i = efeito aleatório do $i^{\text{ésimo}}$ touro;

M_{ij} = efeito aleatório do $j^{\text{ésimo}}$ genótipo do marcador dentro do $i^{\text{ésimo}}$ touro;

e_{ijk} = erro aleatório $\sim N(0, \sigma^2)$.

Sob este modelo, o grupo de progênie que recebeu o alelo \mathbf{M}_1 será comparado com o grupo de progênie que recebeu o alelo \mathbf{M}_2 , dentro da família de cada touro.

Assume-se, neste caso, que os touros sejam heterozigotos, não aparentados, que os acasalamentos foram realizados ao acaso e que:

$$\begin{aligned}\mu &= E(y_{ijk}) \\ S_i &\sim N(0, \sigma_s^2) \\ M_{ij} &\sim N(0, \sigma_m^2) \\ e_{ijk} &\sim N(0, \sigma_e^2).\end{aligned}$$

Tem-se, então, que a covariância entre duas meias-irmãs, que receberam o mesmo alelo do marcador do pai em comum, seja

$$\text{Cov}(y_{ijk}, y_{ijk'}) = \sigma_s^2 + \sigma_m^2$$

Supondo que o marcador \mathbf{M}_1 paterno, esteja ligado a um QTL que contribui com σ_Q^2 para a variância genética aditiva, a covariância acima pode ser escrita como

$$\text{Cov}(y_{ijk}, y_{ijk'}) = \frac{1}{4} \sigma_a^2 + [(1-r)^2 + r^2] \frac{\sigma_Q^2}{2}$$

A covariância entre duas meias-irmãs, que receberam diferentes alelos do marcador do pai em comum, é

$$\text{Cov}(y_{ijk}, y_{ijk'}) = \sigma_s^2$$

Novamente, assumindo-se que o marcador \mathbf{M}_1 paterno esteja ligado a um QTL, esta covariância pode ser escrita

$$\text{Cov}(y_{ijk}, y_{ijk'}) = \frac{1}{4} \sigma_a^2 + [2r(1-r)] \frac{\sigma_Q^2}{2}.$$

O valor do componente σ_m^2 , dentro de touro, será:

$$\text{Cov}(y_{ijk}, y_{ijk'}) - \text{Cov}(y_{ijk}, y_{ij'k'})$$

$$\sigma_m^2 = (1 - 2r)^2 \frac{\sigma_Q^2}{2},$$

Este valor não será nulo, a menos que r seja igual a 0,5. O esquema da análise de variância é apresentado na Tabela 2.

Tabela 2. Esquema de análise de variância em modelo para detecção de QTL ligado a um marcador, em um delineamento de filhas, considerando várias famílias de meias-irmãs (dados balanceados)

Fonte de Variação	GL*	QM	E(QM)	F
Touro	t-1	SQT/(t-1)		
Marcador dentro de touro	(n - t)	SQM:T/(n-t)	$\sigma_e^2 + n\sigma_m^2$	QMM:T/QME
Resíduo	N-n	SQE/(N-n)	σ_e^2	
Total	N-1			

*t=número de touros; n= número de filhas por touro;N = número total de filhas; n = número de observações por genótipo de marcador;

Pode-se verificar que a estatística **F** adequada para verificar a significância do efeito de marcador dentro de touro (M_{ij}) seria

$$F = \frac{QMM:T}{QME}, \text{ associada a } (n - t) \text{ e } N - n \text{ graus de liberdade.}$$

Desta forma, uma estimativa do efeito do QTL, medido por meio de sua variância aditiva, ponderada pela distância entre QTL e marcador, é dada por

$$\sigma_m^2 = (1 - 2r)^2 \frac{\sigma_Q^2}{2} = \frac{E(QMM:T) - E(QME)}{n}.$$

Em que, assumindo-se a normalidade, a hipótese nula a ser testada seria $H_0: \sigma_m^2=0$.

Uma conclusão imediata é que, na rejeição de H_0 , existe um QTL completamente ligado ao marcador, mostrado pela significância do componente de variância entre subgrupos de filhas, recebendo diferentes alelos de seus pais.

Ao estudarem o poder de detecção pela ANOVA, por meio de modelos hierárquicos, alguns autores (SOLLER e GENIZI, 1978, WELLER *et al.*, 1990 e VAN DER BECK, 1995), afirmaram que o aumento do número de filhas por touro é mais eficiente para a melhoria do poder do teste, do que o aumento do número de famílias, além de exigir menor número de animais genotipados. Geralmente, na estimação do poder de detecção, assume-se que o número de observações seja suficientemente grande para incluir pelo menos, uma família informativa.

2.2 - Métodos fundamentados em modelos aleatórios

De acordo com AMOS e ELSTON (1989), métodos robustos, fundamentados em modelos aleatórios, vêm sendo desenvolvidos (HASEMAN e ELSTON, 1972; FULKER e CARDON, 1994), não requerendo a especificação de modelos genéticos para a detecção da ligação.

Os modelos aleatórios se baseiam 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. Assume-se, portanto, que ele seja a proporção média de alelos IBD compartilhados por dois indivíduos, ou em outras palavras, igual ao numerador

do coeficiente de parentesco de Wrigth, 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 IBD em relação ao QTL podem ser diferentes, dado que os indivíduos podem compartilhar ou não, os mesmos alelos. Nesse caso, dois irmãos completos, que compartilham um mesmo alelo do QTL, são geneticamente mais semelhantes do que outros dois irmãos que não compartilham nenhum. Obviamente, como os alelos do QTL não podem ser observados diretamente, essa proporção é tomada a partir dos genótipos observados dos marcadores.

Inicialmente, a utilização de modelos aleatórios foi proposta para a análise de dados de pares de irmãos (HASEMAN e ELSTON, 1972), em estudos de genética humana, onde há predominância de famílias pequenas, de irmãos completos. Esse método se baseia na regressão linear dos quadrados das diferenças fenotípicas entre dois irmãos, dentro de uma família, em função da proporção de IBD compartilhados por eles. Apesar de ser considerado robusto para diversas distribuições dos dados e independência do modelo genético do QTL, ele mostrou ser limitado, pois o efeito genético devido ao QTL e a taxa de recombinação estão confundidos. Posteriormente, foi expandido por GÖTZ e OLLIVIER (1992) para a aplicação em situações típicas do melhoramento animal, onde as possibilidades de estimação dos efeitos do QTL são freqüentemente mais vantajosas do que em humanos. Segundo esses autores, geralmente, as espécies de animais domésticos possuem famílias maiores, informações completas de pedigree e dos marcadores nos pais e nas progênes. Nessa situação, pode-se usar o delineamento de irmãos completos para o caso de grandes famílias de irmãos completos e a análise de meios-irmãos ou delineamento de filhas (NEIMANN-SOESSEN e ROBERSON, 1961), para o caso de grandes famílias de meios-irmãos.

Na tentativa de separar a variância devida ao QTL e o parâmetro de ligação e ainda, localizar o QTL em uma posição específica do cromossomo, FULKER e CARDON (1994) desenvolveram um procedimento de pares de

irmãos, usando o mapeamento por intervalo. Entretanto, apesar do maior poder estatístico, esse método se fundamenta em quadrados mínimos, não otimizando o uso de toda a informação contida nos dados, como as propriedades distribucionais, como faria a máxima verossimilhança (ML).

Segundo MARTINEZ e VUKASINOVIC (2000), o primeiro método para estimar a variância genética devida a determinada região do cromossomo, com base na ML, chamado método de pontos múltiplos de IBD, foi proposto por GOLDGAR (1990). SCHORK (1993) expandiu este método de modo a estimar simultaneamente a variância de várias regiões do cromossomo e o efeito do ambiente comum partilhado por dois irmãos. Embora esses métodos utilizam marcadores flanqueadores para definir o segmento cromossômico, eles não foram elaborados para o mapeamento por intervalo.

Assim, XU e ATCHLEY (1995) propuseram procedimento de mapeamento por intervalo, fundamentado no método de ML desenvolvido por GOLDGAR (1990), onde se assumem a distribuição normal dos valores genotípicos do QTL, o ajustamento do QTL e o efeito poligênico, como efeitos aleatórios para a análise de famílias de irmãos completos. MARTINEZ et al. (1999) estenderam o procedimento para situações onde há prevalência de famílias de meias-irmãs, como em gado de leite, e mostraram que QTL com efeitos relativamente grandes, pode ser detectado com alto poder e localizado com bastante acurácia, especialmente, se grande número de famílias e marcadores polimórficos forem usados.

Embora GOLDGAR (1990) tenha usado dois marcadores flanqueadores para definir um segmento cromossômico e a ML para estimar a variância genética, seu método não foi delineado para o mapeamento por intervalo (XU e ATCHLEY, 1995), sendo seu objetivo, testar se no mínimo, um QTL está localizado em algum lugar na região analisada, sugerindo que ele possa ser usado como um primeiro passo na detecção de QTL. Desse modo, quando uma variação significativa devido ao QTL for encontrada em determinada região do cromossomo, outros métodos poderiam ser usados para encontrar a posição específica do QTL. Caso haja grande número de marcadores distribuídos no

genoma, espera-se que o intervalo flanqueado por dois marcadores adjacentes seja pequeno, da ordem de 10 a 15 cM. Como, de modo geral, existem poucos QTLs distribuídos aleatoriamente pelo genoma, espera-se que a probabilidade de se encontrar dois QTLs em um mesmo intervalo seja reduzida, podendo ser negligenciada. Assim, desenvolveu-se esse método para estender aquele proposto por GOLDGAR (1990) para a situação de mapeamento por intervalo, usando modelos aleatórios para estimar a variância em populações exogâmicas.

XU e ATCHLEY (1995) propuseram dois modelos: o primeiro, com apenas um QTL presente no cromossomo e o segundo, com múltiplos QTL. O modelo com um QTL é semelhante ao de GOLDGAR (1990) e pode ser descrito como:

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

onde

y_{ij} = característica fenotípica medida na $j^{\text{ésima}}$ progênie da $i^{\text{ésima}}$ família;

μ = 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 $\sim N(0, \sigma_e^2)$.

Sob a suposição de equilíbrio de ligação, a variância de y_{ij} é

$$\text{Var}(y_{ij}) = \sigma^2 = \sigma_a^2 + \sigma_g^2 + \sigma_e^2$$

e a covariância entre dois indivíduos não-endogâmicos da mesma família pode ser

$$\text{Cov}(y_{ij}, y_{i'j'}) = \pi_{iq} \sigma_g^2 + \frac{1}{2} \sigma_a^2$$

Deve-se considerar que os π 's são as proporções IBD dos QTL e que elas não são observáveis. XU e ATCHLEY (1995) definiram teoricamente, a covariância condicional dos π 's de todos os marcadores no grupo de ligação e incluíram todos os componentes de variância explicados por cada marcador, na função de verossimilhança, de modo a controlar os efeitos dos outros QTL. Todavia, a técnica não se mostrou eficiente em função do número de parâmetros a serem estimados. ZENG (1993), usando as mesmas propriedades das variáveis indicadoras, pressupôs que a probabilidade condicional do loco de cada marcador flanqueador não seja informativa para o loco do QTL fora do intervalo definido, sendo a única diferença para que as variáveis indicadoras sejam observadas sem erro, ao passo que as proporções IBD dependem do tipo de acasalamento. Assim, assume-se que um marcador seja suficiente para absorver a correlação entre o loco QTL à sua esquerda e o outro à sua direita, tornando-se necessária somente a presença de dois marcadores adicionais (co-fatores), flanqueando o intervalo de interesse, para efetuar essa absorção. Usando a notação **1** e **2**, para denotar os marcadores flanqueando o intervalo de interesse e **L** e **R** para os co-fatores à esquerda e à direita do intervalo, tem-se a disposição **L** – **1** – **2** – **R**, em que o intervalo de interesse está entre os marcadores **1** e **2**. Tomando-se π_{iL} e π_{iR} como os valores IBD dos marcadores à esquerda (**L**) e à direita (**R**), respectivamente, e θ_{kL} e θ_{rR} , como as frações de recombinação entre o $k^{\text{ésimo}}$ QTL e o loco **L** e o $r^{\text{ésimo}}$ QTL com o loco **R**, a covariância condicional entre irmãos completos seria:

$$\text{Cov}(y_{i1}, y_{i2} | \pi_{iL}, \hat{\pi}_{iq}, \pi_{iR}) = \sigma^2 [\pi_{iL} H_L^2 + \hat{\pi}_{iq} h_g^2 + \pi_{iR} H_R^2 + \frac{1}{2} h_a^2]$$

onde

$$H_L^2 = \sum_{k=1}^S (1 - 2\theta_{kL}^2) h_k^2$$

$$H_R^2 = \sum_{r=1}^W (1 - 2\theta_{rR}^2) h_r^2$$

$$h_k^2 = \sigma_k^2 / \sigma^2$$

$$h_r^2 = \sigma_r^2 / \sigma^2$$

Para uma determinada posição, os parâmetros a serem estimados pela função de verossimilhança são $\mu, \sigma^2, H_L^2, h_g^2, H_R^2$ e h_a^2 , mas somente h_g^2 é testada.

Para a obtenção do máximo da função de verossimilhança em relação aos parâmetros, pode ser usado o algoritmo simplex (NELDER e MEAD, 1965).

A hipótese nula é $H_0: h_g^2 = 0$, ou seja, não há QTL segregando no intervalo. A máxima verossimilhança sob H_0 é usada para a obtenção do teste de razão de verossimilhança, que segue a distribuição de Qui-quadrado com $2 \geq GL > 1$ sob H_0 . Um GL é devido ao ajustamento para h_g^2 e os demais, para o ajustamento da posição do QTL, dependendo do tamanho do intervalo. O outro valor é menor do que 2, porque a busca do QTL é realizada dentro de um intervalo testado e não em todo o genoma. Caso a busca seja em todo o genoma coberto pelos marcadores, então $GL \approx 2$ sob H_0 .

3 - MATERIAL E MÉTODOS

As populações e as análises de dados usadas nesse estudo foram simuladas por meio do sistema GeneSibs, que é constituído de três programas (SIM_HS_1Q, IBD_HS_1Q e EST_HS_1Q), escritos para o compilador FORTRAN. Dois desses programas foram escritos pelos doutores Marcos Vinicius Gualberto Barbosa da Silva e Mário Luiz Martinez, ambos da Embrapa Gado de Leite, Juiz de fora – MG. O outro programa, usado para a obtenção das estimativas de máxima verossimilhança, foi adaptado a partir do algoritmo proposto por XU E ATCHLEY (1995).

Este sistema permite a simulação de genomas para estudos relacionados com a detecção de QTL, por meio de modelos aleatórios, podendo simular genomas que possuam até dois QTL. Utiliza o mapeamento por intervalo por meio do método de máxima verossimilhança (XU E ATCHLEY, 1995). O algoritmo é dinâmico e permite a utilização e a comparação de diferentes modelos estatísticos e situações experimentais. Os métodos numéricos implementados incluem procedimentos para mapear QTL, em populações exogâmicas, formadas por meios-irmãos e irmãos completos, usando múltiplos marcadores em uma única característica.

Procedimentos para testar hipóteses relacionadas aos QTL, como o modo de ação gênica, diferentes tamanhos amostrais, herdabilidades da característica, variância devida ao QTL, número de alelos e de marcadores, foram implementados, usando simulação estocástica.

3.1 -SIM_HS_1Q

Este é o programa inicial do sistema GeneSibs destinado à simulação de genótipos e fenótipos de famílias de meios-irmãos. Para tanto, necessita de informação a respeito da estrutura da população e do genoma hipotético, como média, variância, herdabilidade da característica, variância genética devida ao efeito do QTL (em termos percentuais da variância genética total), número de famílias (ou Touros), tamanho de família (número de filhos /touro), percentual de genótipos desconhecidos (de touros e de vacas), número e posição no cromossomo, do QTL e dos marcadores moleculares, número de alelos do QTL e dos marcadores, tamanho do genoma (em cM) e distância entre os locos subsequentes.

Este programa permite alocar o QTL e os marcadores em posições específicas dentro de um cromossomo. Para tanto, o processo de alocação consiste na construção de dois vetores, cada um atuando como um cromossomo homólogo, em que os alelos dos QTL e dos marcadores serão posicionados de acordo com os endereços fornecidos ao sistema.

Segundo SILVA (2003), a partir dos parâmetros iniciais, são estimadas as frequências de recombinação entre os locos, usando-se a função de mapeamento de Haldane e calculadas as variâncias genéticas, dos efeitos dos QTL, os efeitos poligênica e residual, bem como os efeitos de substituição para cada um dos alelos dos QTL. A dominância e outros efeitos aleatórios como, por exemplo, o de ambiente comum, não foram considerados, mas poderão ser incorporados ao modelo posteriormente. Para maior simplicidade, assumiu-se que os valores fenotípicos foram pré-corrigidos para efeitos fixos como rebanho-ano-estação, idade do animal e sistema de manejo.

A geração das famílias tem como base a informação do número de touros e vacas. O programa gera o primeiro touro com base em um número aleatório (n entre 0 e 1), sendo este número transformado em um número inteiro, entre 1 e o número máximo de alelos do marcador (ou QTL). Caso o número gerado seja

maior que o número de alelos , toma-se um novo número e passa-se a gerar o próximo alelo e assim sucessivamente até que todos os alelos sejam gerados.

Como os touros homozigotos não são informativos, o sistema verifica se o animal gerado é homozigoto para um dos locos do marcador. Caso isto seja confirmado, o animal é eliminado e outro touro é gerado.

Após a geração do touro heterozigoto, é calculado o seu efeito genético e o sistema inicia a geração dos genótipos das vacas com as quais este touro é acasalado. Este processo de geração dos genótipos das vacas segue o mesmo procedimento da geração dos touros, não excluindo as vacas homozigotas.

Após a geração dos pais, tem-se a formação dos gametas por um processo denominado “Caminhada Cromossômica”, que consiste em, a partir de uma extremidade de cada cromossomo, iniciar a formação de gametas. Considerando como sendo um passo (do), a distância entre dois locos adjacentes, a caminhada consiste em, a cada passo, simular um número (ds), segundo a distribuição uniforme e no intervalo de 0 a 0,5. Esse valor é, então, transformado em cM pela função de Haldane e comparado com a frequência de recombinação entre os dois locos do passo. Caso o valor simulado e transformado (dr) seja menor que o observado (do), processa-se um “crossing-over” (EUCLYDES, 1996).

A caminhada é executada para cada cromossomo gerando dois gametas, sendo um utilizado para combinar com um outro gameta do outro reprodutor, para a formação da progênie. Cada combinação de gametas entre um touro e uma vaca gera apenas uma progênie, ou seja, é necessária uma caminhada cromossômica para cada acasalamento, sendo possível que os gametas formados pelo pai comum sejam diferentes a cada caminhada.

Em seguida, cada progênie tem o seu valor fenotípico da característica estabelecido, a partir do seguinte modelo:

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

Onde:

y_{ij} = valor fenotípico do $j^{\text{ésimo}}$ indivíduo na $i^{\text{ésima}}$ família de meio-irmão;

μ = média da população;

q_{ij} = efeito dos genótipos dos QTL 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.

Seguindo estes procedimentos, foram simuladas populações considerando um cromossomo de 100 cM, com a presença de um QTL a 50 cM, 6 marcadores espaçados a cada 20 cM, sendo a herdabilidade da característica igual à 0,25, número de famílias igual a 10 e 100, com 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%).

O programa SIM_HS_1Q gera um arquivo contendo o valor fenotípico da progênie, informações de pedigree e genótipos do pai, da mãe e da progênie.

3.2 - IBD_HS_1Q

O próximo passo após a geração das famílias é o cálculo da proporção de genes idênticos por descendência (IBD), realizado por este programa, que gera a matriz de proporção de genes idênticos por descendência, baseado no algoritmo proposto por MARTINEZ E VUKASINOVIC (2000).

Assumindo-se que \mathbf{a} represente qualquer par de meios-irmãos, a estimativa da proporção de IBD ($\hat{\pi}_a$) entre dois meios-irmãos será dada, segundo MARTINEZ e VUKASINOVIC (2000), por:

$$\hat{\pi}_a = \frac{1}{2} \cdot f_a$$

em que f_a é a probabilidade que o par de meios-irmãos \mathbf{a} tenha um IBD em um loco ao acaso.

MARTINEZ e VUKASINOVIC (2000) afirmaram que, como as famílias de meios-irmãos têm apenas um pai em comum, o número possível de combinações de acasalamento é igual $[n(n+1)/2]^3$, em que n é o número de alelos em um determinado loco, sendo o número de genótipos possível igual a $(n+1)/2$, pois existem três indivíduos (um pai e duas mães) envolvidos na produção de um determinado par de meios-irmãos

A estimação de π quando ambos os pais e as progênes têm o genótipo conhecido, considerando que o acasalamento de machos e fêmeas ocorra sem nenhum parentesco entre si (entre os pais e as mães, entre as mães e entre os pais) e que, em determinado loco, os alelos tenham a mesma frequência e assumindo-se, ainda, que todos os pais sejam heterozigotos para os locos dos marcadores, a estimativa de π_a ($\hat{\pi}_a$), segundo MARTINEZ e VUKASINOVIC (2000), é conseguida pela seguinte maneira:

1º) Obtenção de todos os possíveis genótipos das progênes resultantes dos acasalamentos de cada pai com todas as mães;

2º) Determinação dos tipos para todas as combinações de pares de irmãos;

3º) Determinação do número de alelos IBD entre pares de irmãos. para o caso de meio-irmãos, quando os pais não são consangüíneos, o número de alelos IBD será 1 (se ambos os irmãos o mesmo alelo do pai comum) ou zero (se os irmãos tiverem recebido alelos diferentes do pai comum);

4º) Cálculo da probabilidade (f_a) de que determinado par de irmãos **a** tenha um alelo IBD no loco em questão. para se calcular f_a , usa-se:

$$f_a = \frac{N_{IBD}}{N_t}$$

em que N_{IBD} é o número de pares de irmãos do tipo **a**, que têm um alelo ibd, e N_t é o número total de pares de irmãos do tipo **a**.

5º) Cálculo da proporção de alelos IBD ($\hat{\pi}_a$) para determinado par de irmão do tipo a.

Neste mesmo trabalho, MARTINEZ e VUKASINOVIC (2000) apresentam os procedimentos para se estimar π , quando apenas o pai em comum e as progênes têm o genótipo conhecido e também quando apenas as progênes têm o genótipo conhecido

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 os mesmos alelos do QTL.

A equação para a obtenção da estimativa de π_q é

$$\hat{\pi}_q = E(\pi_q | \pi_1 \pi_2) = \alpha + \beta_1 \pi_1 + \beta_2 \pi_2$$

onde, 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:

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

Definidas as taxas de recombinação θ_1 , θ_2 e θ_{12} , e substituindo-se $\bar{\pi} = 1/4$,

$V(\pi) = 1/16$ e $\text{Cov}(\pi_i, \pi_j) = \frac{(1 - 2\theta_{ij})^2}{16}$, que são os resultados apresentados por MARTINEZ e VUKASINOVIC (1999) para meio-irmãos, os coeficientes podem ser estimados como:

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

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

$$\hat{\alpha} = (1 - \hat{\beta}_{\pi_1} - \hat{\beta}_{\pi_2}) / 4.$$

3.3 - EST_HS_1Q

Uma vez gerados os genomas das famílias, os valores fenotípicos e calculada a proporção de alelos IBD, o próximo passo é a estimação de parâmetros, usando a máxima verossimilhança no mapeamento por intervalo.

O modelo usado é o mesmo proposto por GOLDGAR (1990).

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

Onde:

Y_{ij} = característica fenotípica medida na $j^{\text{ésima}}$ progênie da $i^{\text{ésima}}$ família;

μ = 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 $\sim N(0, \sigma_e^2)$.

Assume-se que todos os efeitos aleatórios do modelo sigam a distribuição normal. Entretanto, se σ_a^2 e σ_e^2 forem suficientemente grandes para tornar os dados normalmente distribuídos, não é exigida a distribuição normal do efeito do QTL.

Para uso desse modelo, pressupõe-se que haja ausência de interferência e de dominância, seguindo o número de crossing-overs distribuição de Poisson,

ocorrendo à distancia de λ Morgans. Também se assume que os crossing-overs ocorrendo em determinado intervalo sejam uniformes e independentemente distribuídos na região analisada.

A variância de y_{ij} , assumindo equilíbrio de ligação é:

$$\text{Var}(y_{ij}) = \sigma^2 = \sigma_a^2 + \sigma_g^2 + \sigma_e^2$$

A covariância entre dois indivíduos não endogâmicos da mesma família é:

$$\text{Cov}(y_{ij}, y_{ij'}) = \pi_q \sigma_q^2 + 1/4 \sigma_a^2$$

Onde π é a proporção de alelos IBD do suposto QTL compartilhado pelos irmãos.

Já o coeficiente associado à variância poligênica é $1/4$ pois, em média, espera-se que dois meios-irmãos compartilhem esta proporção de IBD. Para as proporções de IBD do QTL, o valor pode variar entre 0 e $1/2$, dependendo do par de irmãos.

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

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

onde

$$C_i = \begin{pmatrix} 1 & r_i \\ r_i & 1 \end{pmatrix}$$

e

$$r_i = \hat{\pi}_q h_g^2 + 1/4 h_a^2$$

Se cada família possuir k meio-irmãos, sendo C_i é uma matriz $k \times k$, as herdabilidades são definidas como:

- $h_g^2 = \frac{\sigma_g^2}{\sigma^2}$ - herdabilidade do suposto QTL;
- $h_a^2 = \frac{\sigma_a^2}{\sigma^2}$ - herdabilidade do componente poligênico, e
- $h_t^2 = \frac{(\sigma_g^2 + \sigma_a^2)}{\sigma^2}$ - herdabilidade total

Assim, assumindo-se que todos os dados têm distribuição normal multivariada, a função densidade conjunta das observações dentro das famílias é:

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

onde $y_i = [y_{i1}, y_{i2}, y_{i3}, \dots, y_{ik}]'$ é um vetor $k \times 1$ de valores fenotípicos observados para k irmãos dentro da família i , e $1 =$ vetor $k \times 1$, com todos os elementos iguais a 1. De acordo com XU e ATCHLEY (1995) a distribuição normal do efeito devido ao QTL não é necessária, dado que sua variância é pequena em relação à soma das variâncias poligênicas e das variâncias de ambiente. Todavia, exige-se a normalidade de a_{ij} e e_{ij} para que y_{ij} tenha distribuição normal.

O log da verossimilhança para n famílias independentes é

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

A função de verossimilhança relaciona a posição do QTL flanqueado pelos dois marcadores por meio de r_i . Os parâmetros desconhecidos a serem estimados são μ , σ^2 , h_g^2 , h_a^2 e θ_{1q} . Para se maximizar 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, vai-se aumentando seu valor e, conseqüentemente, diminuindo-se a distância entre QTL e o marcador posicionado do lado direito (θ_{2q}), em todo o intervalo. Repete-se o procedimento até que todo o genoma tenha sido coberto. A estimativa 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.

Segundo SILVA (2003), em uma determinada posição no intervalo, o algoritmo empregado assume valores *a priori*, de h_g^2 e h_a^2 , obtendo-se estimativas de ML para μ e σ^2 , expressas como funções de h_g^2 e h_a^2 :

$$\hat{\mu} = u(h_g^2, h_a^2) = \left[\sum_{i=1}^n \mathbf{1}' \mathbf{C}_i^{-1} \mathbf{1} \right]^{-1} \left[\sum_{i=1}^n \mathbf{1}' \mathbf{C}_i^{-1} y_i \right]$$

e

$$\hat{\sigma}^2 = v(h_g^2, h_a^2) = \frac{1}{nk} \sum_{i=1}^n (y_i - \mathbf{1}\hat{\mu})' \mathbf{C}_i^{-1} (y_i - \mathbf{1}\hat{\mu})$$

Se houver variação no tamanho das famílias, $n\mathbf{k}$ deve ser substituído por $\sum_{i=1}^n k_i$. Essas duas equações são obtidas, tomando-se a derivada primeira de \mathbf{L} em relação a σ^2 , igualando-a a zero. Então, substitui-se μ e σ^2 na função de verossimilhança original por suas estimativas de ML, sendo o log de verossimilhança expresso como função de h_g^2 e h_a^2 . Dessa maneira:

$$L = -\frac{1}{2}nk \log(\sigma^2) - \frac{1}{2} \sum_{i=1}^n \log(|\mathbf{C}_i|)$$

Nota-se que após a substituição, μ e σ^2 são absorvidos na função de verossimilhança, sendo esta última maximizada em relação a h_g^2 e h_a^2 , via algoritmo simplex (NELDER e MEAD, 1965). Ainda, segundo XU e ATCHLEY (1995), a substituição de μ e σ^2 por \mathbf{u} e \mathbf{v} implica em atualizar as estimativas de h_g^2 e h_a^2 , sendo as funções $\hat{\mu} = u(h_g^2, h_a^2)$ e $\hat{\sigma}^2 = v(h_g^2, h_a^2)$ atualizadas.

A hipótese de nulidade a ser testada é $H_0: h_g^2 = 0$, ou seja, não existe QTL segregando no intervalo testado. A ML sob a hipótese de nulidade é denotada por L_0 . O teste de razão de verossimilhança (LR) é dado por:

$$LR = -2(L_0 - L),$$

que segue a distribuição de Qui-quadrado com $2 \geq GL > 1$ sob H_0 .

O cromossomo vai sendo varrido da esquerda para a direita, a partir de um marcador para o próximo, em passos de 2 cM, sendo LR calculado a cada passo. A posição com a maior razão de verossimilhança é assumida como a posição mais provável do QTL. De modo análogo, as estimativas dos componentes de variância nessa posição são aceitas como as maiores estimativas de ML para os parâmetros h_g^2 e h_a^2 .

Os valores empíricos de limiar para tamanho de famílias igual a 25 foram os mesmos para todos os testes, independentemente do número de alelos do QTL, sendo o 95º percentil igual a 9,2776. Para tamanho de famílias igual a 50, com número de alelos do QTL igual a 2 ou 4, este percentil foi de 5,8693. Para famílias com 50 filhas e número de alelos do QTL igual a 6 ou 10, este percentil foi de 4,4852. Quando havia 100 famílias com 50 filhos este valor foi de 6,9027.

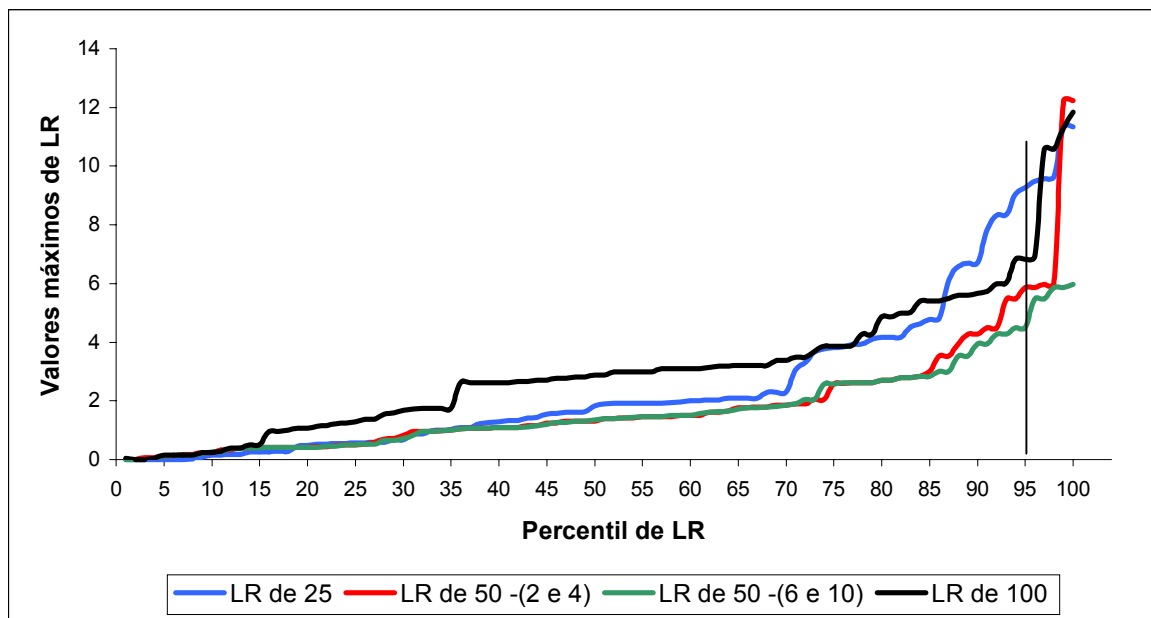


figura 01- Valores máximos do teste de razão de verossimilhança (LR) sob H_0 , para diferentes tamanhos de famílias e número de alelos do QTL.

4 - RESULTADOS E DISCUSSÃO

As estimativas para as localizações dos QTL, relativas às médias de 100 repetições, com o correspondente intervalo de confiança para os diferentes números de famílias, tamanhos de família, número de alelos do QTL e proporções de variância genética devida aos QTL, são apresentadas na tabela 3

Tabela 3. Efeito do número de alelos do QTL, número e tamanho da família e proporção da variância genética devida ao QTL, sobre a estimativa da posição e intervalos de confiança.

NF	TF	N. QTL	Proporção da variância genética devida ao QTL (%)					
			10		50		100	
			cM	IC	cM	IC	cM	IC
10	25	2	46,38	12,55	53,88	9,81	56,16	7,90
		4	47,28	12,30	50,36	10,45	53,24	8,99
		6	48,48	12,28	49,92	10,33	53,50	8,22
		10	48,28	12,25	50,76	9,98	53,40	8,62
10	50	2	59,36	12,21	56,30	12,14	52,90	11,37
		4	61,20	12,44	55,80	13,24	54,26	11,43
		6	59,60	12,91	55,40	13,19	54,60	11,27
		10	62,36	12,49	57,54	12,50	54,82	11,57
100	50	4	56,8	11,14	52,62	5,95	50,3	2,4

TF = Tamanho da família; N.QTL = número de alelos do QTL; cM = local estimado do QTL; IC = intervalo de confiança. NF = número de famílias

O intervalo de confiança é calculado como sendo igual a 4 vezes o erro-padrão.

Para 10 famílias com 25 filhos, o aumento no número de alelos do QTL teve pouca influência sobre sua localização. A melhor estimativa foi quando se tinha 6 alelos no QTL e 50 % da variância genética era devida ao mesmo. Quando havia 10 % da variância genética devida ao QTL, as estimativas de localização foram subestimadas (46,38 à 48,28), havendo o aparecimento de uma imagem fantasma, sugerindo a existência de um QTL próximo à posição 80 cM, como pode ser observado nas figuras de 2 a 5. Com 50 % da variância genética, as estimativas de localização (49,92 à 53,88) são as mais próximas do valor real (50), principalmente para 4, 6 ou 10 alelos do QTL. Houve o surgimento de uma imagem fantasma, sugerindo a existência de um outro QTL na posição 70 cM, principalmente quando existam 2 alelos. Já para 100% da variância genética devida ao QTL, as estimativas de localização são superestimadas (53,24 a 56,16), não havendo surgimento de imagem fantasmas e sim, dominância de um único pico de LR na posição 50 cM, embora o valor de LR nesta situação não tenha excedido a 4,5, ficando abaixo do valor empírico de limiar (9,2776)

De um modo geral, o número de alelos do QTL não influenciou nas estimativas dos intervalos de confiança. Houve pequenas variações dentro de cada combinação de tamanho de família e variância genética. No aumento da proporção da variância genética devida ao QTL de 10 para 100%, houve uma diminuição no intervalo de confiança das estimativas de localização, da ordem de 30% para tamanho de família igual a 25 e de 10% para tamanho de família igual a 50.

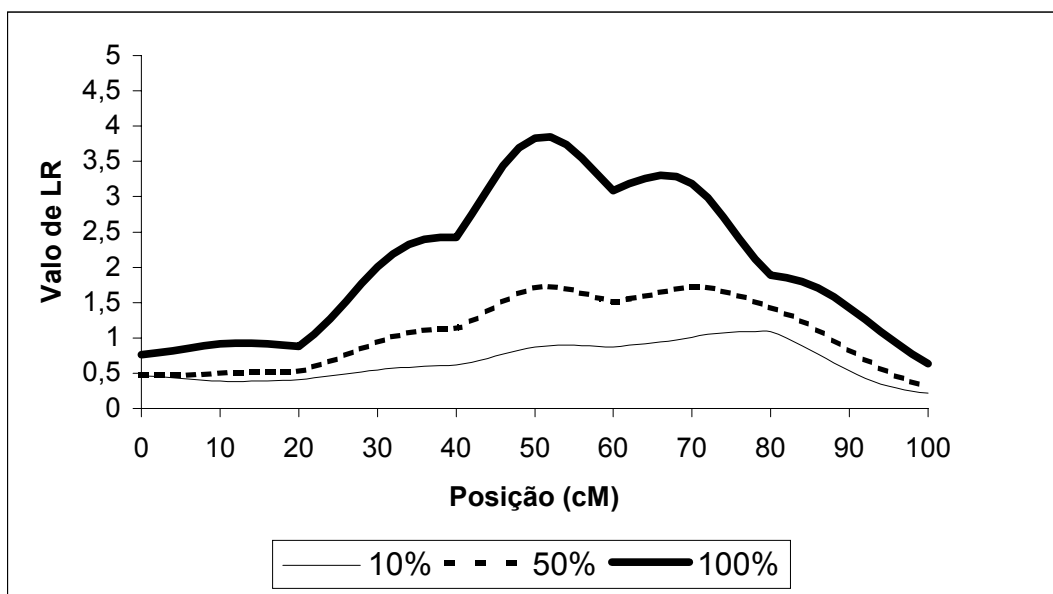


Figura 2- Comparação dos valores de LR para três diferentes proporções de variâncias genéticas devidas aos QTL (10,50 e 100%), para número de família = 10 e tamanho de família = 25, com número de alelos do QTL = 2

Figura 3- Comparação dos valores de LR para três diferentes proporções de variâncias genéticas devidas aos QTL (10,50 e 100%), para número de família = 10 e tamanho de família = 25, com número de alelos do QTL = 4

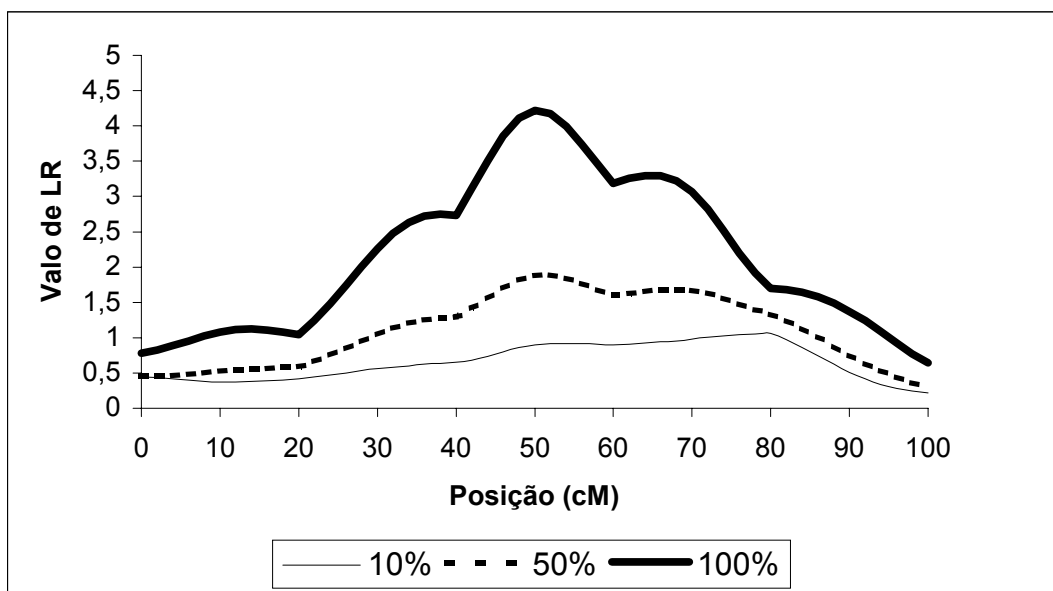


Figura 4- Comparação dos valores de LR para três diferentes proporções de variâncias genéticas devidas aos QTL (10,50 e 100%), para número de família = 10 e tamanho de família = 25, com número de alelos do QTL = 6

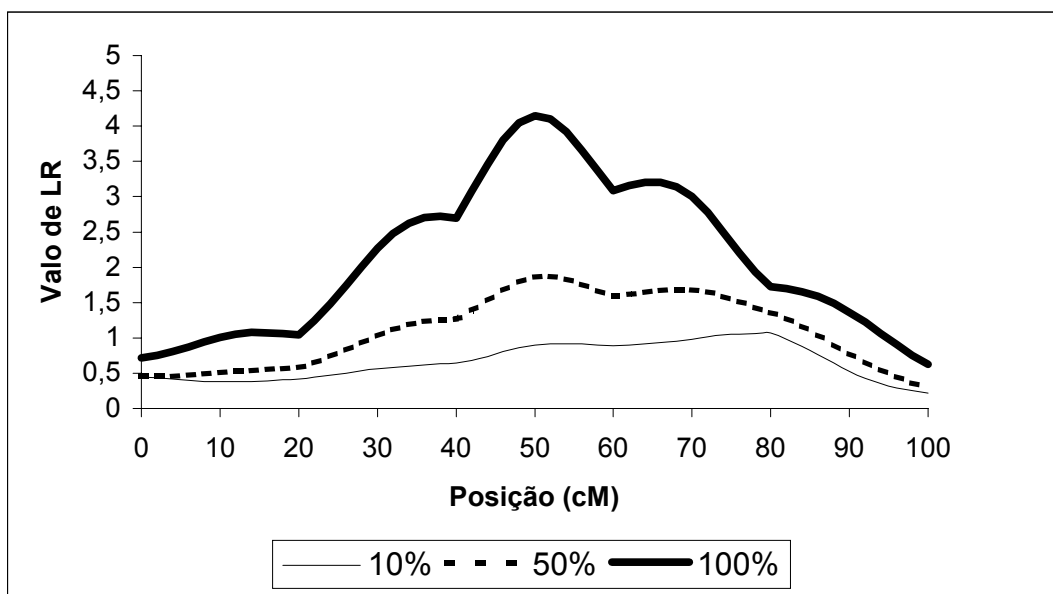


Figura 5- Comparação dos valores de LR para três diferentes proporções de variâncias genéticas devidas aos QTL (10,50 e 100%), para número de família = 10 e tamanho de família = 25, com número de alelos do QTL = 10

Para número de famílias igual a 10 e tamanho de família igual a 50, o número de alelos do QTL também não influenciou na localização do mesmo, sendo o melhor resultado, com 2 alelos do QTL e 100% da variância genética devida ao QTL. Com 10% da variância genética, houve uma flutuação ao longo de todo o genoma, como pode ser observado nas figuras de 6 a 9. Com 50% da variância genética, observa-se uma flutuação de valores de LR por todo o cromossomo com uma tendência de localizar um QTL entre as posições 40 e 50 cM. Já com 100% da variância genética devido ao QTL, há uma pequena predominância de um pico de LR na posição 50 cM, com tendência de se formarem imagens fantasmas em ambos os lados, mas principalmente à esquerda. Houve uma aproximação dos valores do gráfico em relação aos valores empíricos de limiar (5,8693 para 2 ou 4 alelos no QTL, e 4,4852 para 6 ou 10 alelos no QTL), tendo todos os valores permanecido abaixo do valor de limiar.

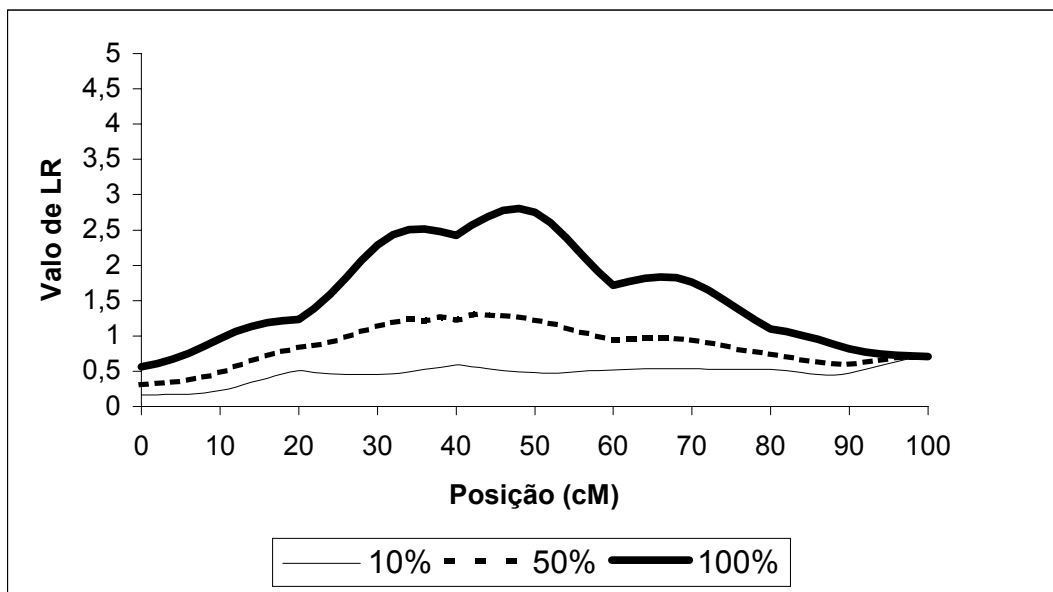


Figura 6- Comparação dos valores de LR para três diferentes proporções de variâncias genéticas devidas aos QTL (10,50 e 100%), para número de família = 10 e tamanho de família = 50, com número de alelos do QTL = 2

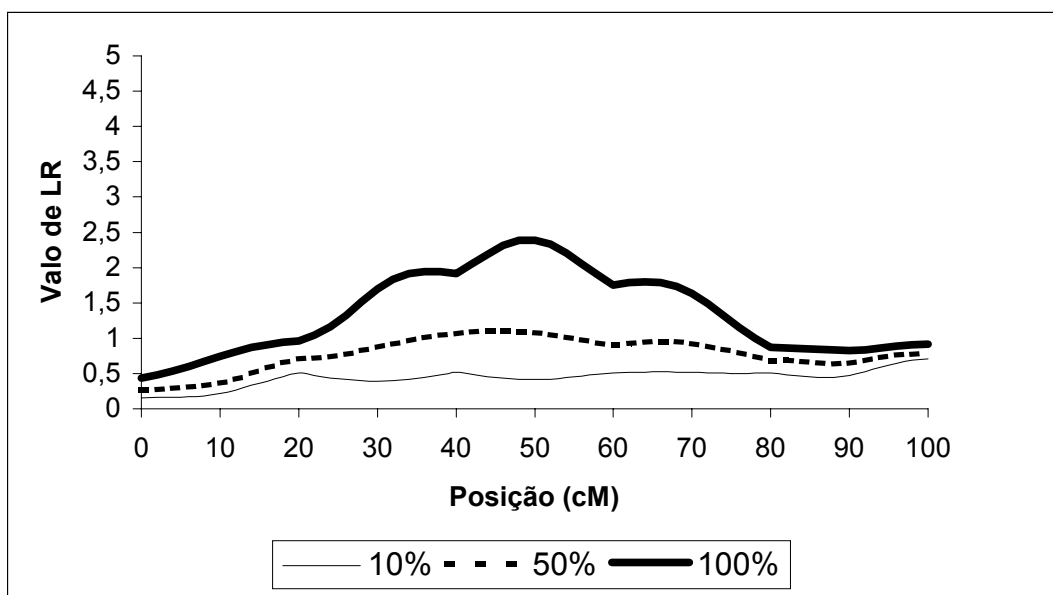


Figura 7- Comparação dos valores de LR para três diferentes proporções de variâncias genéticas devidas aos QTL (10,50 e 100%), para número de família = 10 e tamanho de família = 50, com número de alelos do QTL = 4

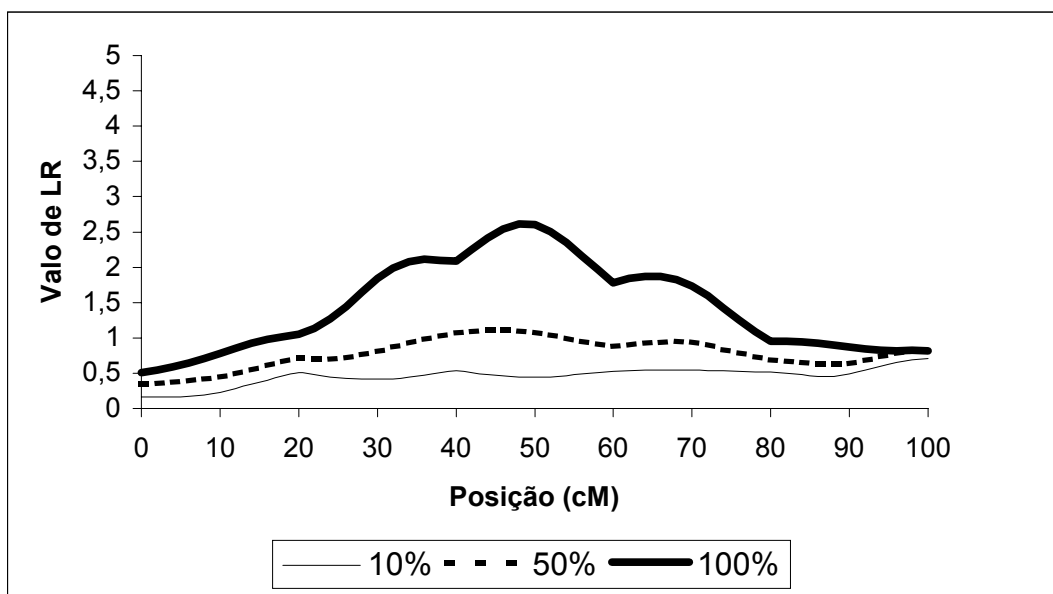


Figura 8- Comparação dos valores de LR para três diferentes proporções de variâncias genéticas devidas aos QTL (10,50 e 100%), para número de família = 10 e tamanho de família = 50, com número de alelos do QTL = 6

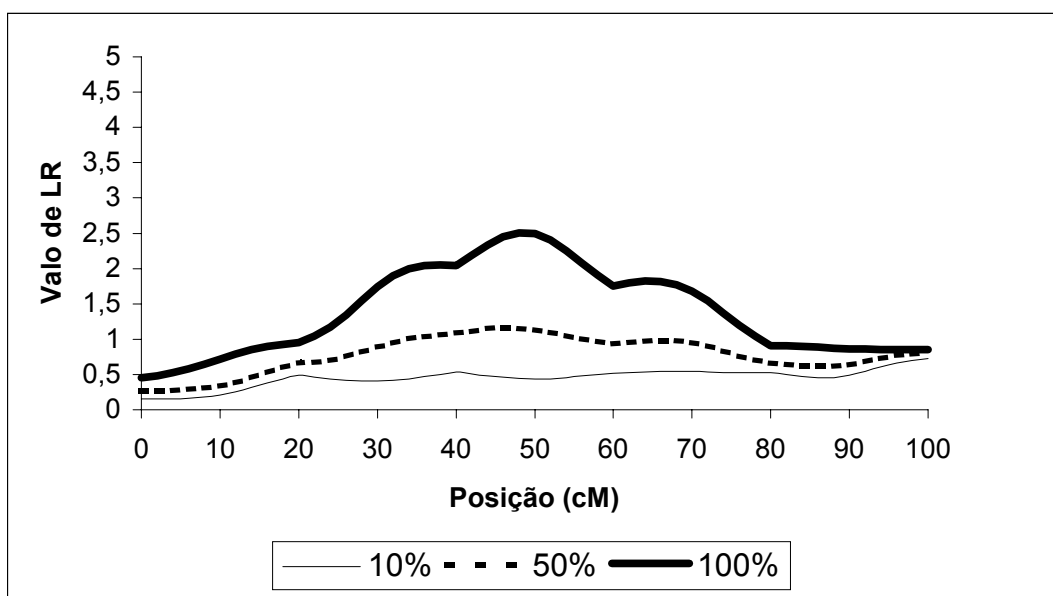


Figura 9- Comparação dos valores de LR para três diferentes proporções de variâncias genéticas devidas aos QTL (10,50 e 100%), para número de família = 10 e tamanho de família = 50, com número de alelos do QTL = 10

Quando se tinha baixa proporção da variância genética devida ao QTL (10%), houve a tendência de indicar erroneamente a presença de QTLs, independentemente do tamanho de família e número de alelos do mesmo. Uma

possível explicação para este fato, segundo MARTINEZ et al. (1999), pode ser a baixa herdabilidade da característica, aliada à reduzida variação genética explicada pelo QTL e ao número de famílias usadas, tornando as estimativas viesadas.

Quando se aumentou o número de famílias para 100, com 50 filhas cada e 100% da variação devida ao QTL, houve localização quase exata do QTL (50,3) com um intervalo de confiança baixo (2,4) com predominância de um pico de LR na posição 50 cM, como pode ser observado na figura 10. O aumento no número de famílias melhorou as estimativas de localização e contribuiu para a redução dos valores dos intervalos de confiança, melhorando a acurácia das estimativas, indicando como uma das possíveis causas para o viés observado, o reduzido número de famílias utilizadas.

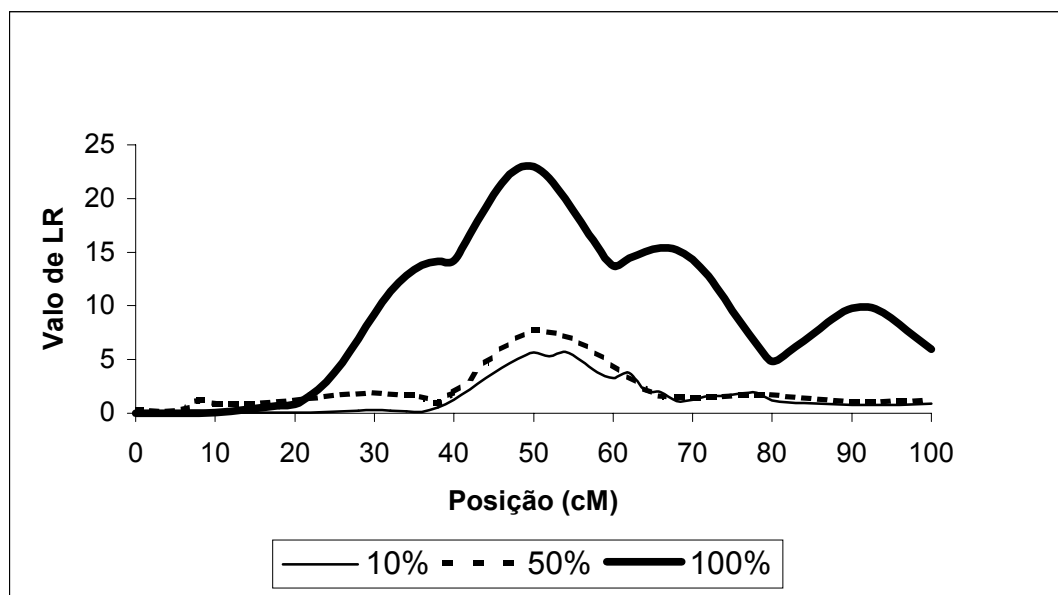


Figura 10- Comparação dos valores de LR para três diferentes proporções de variâncias genéticas devidas aos QTL (10,50 e 100%), para número de família = 100 e tamanho de família = 50, com número de alelos do QTL = 4

As estimativas dos componentes de variância, apresentadas como herdabilidade do QTL (h_q^2), herdabilidade poligênica (h_a^2) e herdabilidade total (h_t^2) estão na tabela 4. Sendo que as estimativas de h_t^2 correspondem às somas de h_q^2 e h_a^2 .

Para tamanho de família igual a 50, as estimativas de h_t^2 foram todas subestimadas, não tendo havido diferença nas estimativas com a variação no

número de alelos do QTL. Quando se aumentou a proporção da variância genética devido ao QTL houve uma aproximação do valor estimado com o valor verdadeiro (0,25), porém todas as estimativas foram subestimadas. O valor mais próximo do real ocorreu com o número de alelos do QTL igual a 2 e com toda a variância genética devido ao QTL.

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

Tabela 4. Efeito do número de alelos do QTL, tamanho da família e proporção da variância genética devida ao QTL, sobre as estimativas de herdabilidade devidas ao QTL (h_q^2), poligênica (h_a^2) e total (h_t^2).

NF	TF	N.QTL	Proporção da variância genética devida 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	0,01	0,16	0,17	0,06	0,16	0,22	0,08	0,17	0,25

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

SILVA (2003), trabalhando com número de famílias igual a 50, obteve estimativas de h_t^2 muito próximas do valor paramétrico verdadeiro, mesmo quando a variação devido ao QTL foi de apenas 20%. Estes resultados de SILVA (2003) foram consistentes com os obtidos por MARTINEZ et al. (1998), que revelam terem sido obtidas estimativas com maior acurácia, quando se aumentou o número de animais genotipados.

Quando havia 100 famílias com 50 filhas cada e 100% da variância genética devido ao QTL, o valor da estimativa de h_t^2 foi igual ao valor paramétrico verdadeiro (0,25), confirmando a afirmação de MARTINEZ et al. (1998).

Mesmo quando as estimativas de h_t^2 foram iguais ou próximas do valor verdadeiro, não foram obtidas separações adequadas desses valores em seus componentes, devidas ao QTL (h_q^2) e ao efeito poligênico (h_a^2). De acordo com MARTINEZ et al. (1999), a separação desses componentes pode ser eficientemente obtida, somente quando a variação devida ao QTL explica aproximadamente 10-15% da variância genética. Segundo esses autores, as variâncias de QTL de pequeno efeito tendem a ser superestimadas, enquanto as variâncias de QTL de grandes efeitos são sempre subestimadas, com o aumento do viés a cada elevação da variância devida ao QTL. Entretanto, pode ser observado também, que a subestimação da variância do QTL é sempre acompanhada de uma superestimação do efeito poligênico, conservando a soma de h_q^2 e h_a^2 próxima do valor verdadeiro, podendo indicar uma eficiente separação das variâncias genética e residual (MARTINEZ et al., 1999). GESSLER e 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 , foi igual a zero, independentemente da variância devida ao QTL, do tamanho da família e do número de alelos do QTL. Os mesmos valores foram obtidos por SILVA (2003), quando o número de

famílias foi igual a 10, afirmando que o poder de detecção foi altamente dependente da proporção da variância genética devida 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 com 40 indivíduos cada, em que o QTL era responsável por toda a variância genética.

MARTINEZ et al. (1999) afirmaram que o poder de detecção do QTL em geral, é altamente dependente da herdabilidade da característica e da proporção da variância genética explicada por ele. 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 contribuem para o poder de detecção.

5 – CONCLUSÕES

O mapeamento por intervalo baseado em modelos aleatórios em populações de meios-irmãos para a identificação de QTL não foi apropriado para investigar a presença de QTL, quando havia apenas 10 famílias com 25 ou 50 filhos. Todavia, quando se aumenta o número de famílias e o número de filhos por família, o mapeamento do QTL é bem preciso e acurado.

O número de alelos do QTL não influenciou na estimação da localização ou na estimação dos componentes de variância.

O método indicou corretamente o intervalo do suposto QTL.

Recomenda-se o modelo aleatório para a varredura rápida do genoma, associado a análises mais refinadas no intervalo em que for indicada a existência de um suposto QTL.

O modelo aleatório não separou eficientemente as variâncias poligênicas e dos QTLs, devendo ser usados métodos mais eficientes para estimar essa última.

6 - REFERÊNCIAS BIBLIOGRÁFICAS

- AMOS, C.I., ELSTON, R.C. 1989. Robust methods for detection of genetic linkage for quantitative data from pedigrees. *Genet. Epidemiol.*, 6: 349-360.
- BINK, M., VAN ARENDONK, J.A.M. 2000. Detection of quantitative trait loci in outbred population with incomplete marker data. *Genetics*, 151:409-420.
- BINK, M., VAN ARENDONK, J.A.M., QUAAS, R.L. 1998. Breeding value estimating with incomplete marker data. *Gen. Sel. Evol.*, 30:45-58.
- BLACKWELDER, W.C., ESLTON, R.C. 1982. Power and robustness of sib-pair linkage tests and extension to larger sibships. *Comum. Statist. Theor. Methods*, 11: 449-484.
- BOILARD, M., BOICHARD, D. 2002. Use of maternal information for QTL detection in a (grand)daughter design. *Genet. Sel. Evol.*, 34:335-352.
- BOVENHUIS, H., VAN ARENDONK, J.A.M., DAVIS, G., et al. 1997. Detection and mapping of quantitative trait loci in farm animals. *Livert. Prod. Sci.*, 52:135-144.
- EUCLYDES, R.F. 1996. Uso do sistema de simulação Genesys na avaliação de métodos de seleção clássicos e associados a marcadores moleculares. Viçosa: UFV, 1996, 150p. Tese (Doutorado em Genética e Melhoramento) – Universidade Federal de Viçosa, MG, 1996.
- FALCONER, D.S., MACKAY, T.F.C. 1996. Introduction to quantitative genetics. 4th ed., Longman, Harlow, UK.

- FERNANDO, R., GROSSMAN, M. 1989. Marker assisted selection using best linear unbiased prediction. *Gen. Sel. Evol.*, 21:467-477.
- FULKER, D.W, CARDON, L.R. 1994. A sib-pair approach to interval mapping of quantitative trait loci. *Am. J. Hum Genet.* , 47: 957-967.
- GESSLER, D.D.G., XU, S. 1996. using the expectation or the distribution of the identity by descent for mapping quantitative trait loci under the radon model. *Am. J. Hum. Genet.*, 59:1382-1390.
- GODDARD, M.E. 1992. A mixed model for analyses of data multiple genetic markers. *Theor. Appl. Genet.*, 83:878-886.
- GOLDGAR, D.E. 1990. Multipoint analysis of human quantitative genetic variation. *Am. J. Hum. Genet.*, 47:957:967.
- GÖTZ, K.U., OLLIVIER, L. 1992. Theoretical aspects of applying sib-pai linkage tests to livestock species. *Genet. Sel. Evol.*, 24:29-42.
- HALDANE, J.B.S. 1919. The combination of linkage values and the calculation of distance between the loci linked factors. *J. Genet.*, 8:299-309.
- HALEY, C.S. , KNOTT, S.A. 1992. A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity*, 69: 315-324.
- HASEMAN, J.K., ELSTON, R.C. 1972. The investigation of linkage between a quantitative trait and a marker locus. *Behav. Genet.*, 2: 3-19.
- HENDERSON, C. R. 1976. A simple method for computing the inverse of numerator matrix used in prediction of breeding values. *Biometrics*, 32:69-83.
- HOESCHELE, I., UIMARI, P., GRIGNOLA, F.E., ZHANG, Q., GAGE, K.M. 1997. Advances in statistical methods to map quantitative trait loci in outbreed populations. *Genetics*, 147:1445-1457.

KADARMIDDEN, H.J., DEKKERS, J.C.M. 1999. Regression on markers with uncertain allele transmission for QTL mapping in half-sibs design. *Genet. Sel. Evol.*, 31:437-455.

LYNCH, M., WALSH, B. 1998. Genetics and analysis of quantitative traits. 1st ed. , Ed. Sinauer Associates, Inc., Sunderland, MA.

MARTINEZ, M. L. 1998. Técnicas de biologia molecular aplicadas à produção animal. São Carlos, SP. 127 p. (Mimeo.)

MARTINEZ, M.L., VUKASINOVIC, N., FREEMAN, A.E. 1998. Estimating QTL location and QTL variance in half-sibs families under the random model with missing parental genotypes. *Anim. Breed. Genet.*, 115:165-180.

MARTINEZ, M.L., VUKASINOVIC, N., FREEMAN, A. E. 1999. Random model approach for QTL mapping in half-sib families. *Genet. Sel. Evol.*, 31:319-340.

MARTINEZ, M.L., VUKASINOVIC, N. 2000. Algoritmo para cálculo da proporção de gens idênticos por descendência, para mapear QTL em famílias de meio-irmãos. *Rev. Bras. Zoot.*, 29:443-451.

NELDER, J.A., MEAD, R. 1965. A simplex method for function minimization. *Comput. J.*, 7:308-313.

NEIMANN-SORENSEN, A., ROBERTSON, A. 1961. The association between blood groups and several production characteristics in three Danish cattle breeds. *Acta Agric. Scand.*, 11:163-196.

SAS Institute Inc. 1990. SAS/STAT *User's guide, Version 6*. SAS Institute Inc, Cary, NC, 4th edition, vol. 1, 846p.

SCHORK, N.J. 1993. Extended multipoint identity-by-descent analysis of human quantitative traits: efficiency, power, and modeling considerations. *Am. J. Hum. Genet.*, 53:1306-1393.

SILVA, M.V.G.B.; 2003. Utilização de modelos aleatórios no mapeamento de QTL em famílias de meio-irmãos e irmãos completos. Viçosa: UFV, 2003. Tese (Doutorado em Genética e Melhoramento) – Universidade Federal de Viçosa, MG, 2003. No prelo.

SOLLER, M., GENIZI, A. 1978. The efficiency of experimental designs for the detection of linkage between a marker locus and a locus affecting a quantitative trait in segregating populations. *Biometrics*, 34:47-55.

VAN ARENDONK, J.A.M., TIER, B., KINGHORN, B.P. 1994. Use of multiple genetic markers in prediction of breeding values. *Genetics*, 137:319-329.

VAN DER BERCK, S., VAN ARENDONK, J.A.M., GROEN, A.F. 1995. Power of two- and three-generation QTL mapping experiments in an outbred populations containing full-sibs or half-sib families. *Theor. Appl. Genet.*, 91:1115-1124.

XU, S. , ATCHLEY, W.R. 1995. A random model approach to interval mapping of quantitative trait loci. *Genetics*, 141: 1189-1197.

WELLER, J.I., KASHI, Y., SOLLER, M. 1990. Power of daughter and granddaughter designs for determining linkage between marker loci and quantitative trait loci in dairy cattle. *J. Dairy. Sci.*, 73:2525-2537.

ZENG, Z.-B. 1993. Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci. *Proc. Natl. Acad. Sci. USA*, 90:10972-10976.