

# IV Curso de Técnicas de Biologia Molecular Aplicadas à Produção Animal

## ANAIS



**30/06 a 04/07/2003**

**Embrapa Pecuária Sudeste**

**Embrapa Pecuária Sudeste**  
Rod. Washington Luiz, km 234 -Fazenda Canchim  
C.Postal, 339 - São Carlos, SP - 13560-970  
Tel.:(0XX16) 261 5611 - Fax:(0XX16) 261 5754  
Email: [sac@cppse.embrapa.br](mailto:sac@cppse.embrapa.br)  
Home Page: [www.cppse.embrapa.br](http://www.cppse.embrapa.br)



## **IV Curso de Técnicas de Biologia Molecular Aplicadas à Produção Animal**

### **Comissão Organizadora:**

Dra. Luciana Correia de Almeida Regitano - Embrapa Pecuária Sudeste - Coordenadora

Dra. Márcia Cristina de Sena Oliveira - Embrapa Pecuária Sudeste - Coordenadora

Prof. Dr. Luiz Lehmann Coutinho - ESALQ/USP - Coordenador

Gilberto Cesar Agostinho - Embrapa Pecuária Sudeste - Membro

Jorge Reti - Embrapa Pecuária Sudeste - Membro

Andréa Shibata de Barros - Embrapa Pecuária Sudeste - Membro

Carlos Policarpo - Embrapa Pecuária Sudeste - Membro

Carlos Roberto de Souza Paino - Embrapa Pecuária Sudeste - Colaborador

Emília Maria Pulcinelli Camarnado - Embrapa Pecuária Sudeste - Membro

Silmara Perez Barcellos - Embrapa Pecuária Sudeste - Membro

Maria Cristina Campanelli Brito - Editoração

### **Patrocinadores:**

-Biosystems

- Comercial, Importadora, Exportadora de Equipamentos para Laboratórios Ltda.

-Amersham Bioscience

**Embrapa Pecuária Sudeste**  
**Rod. Washington Luiz, km 234 -Fazenda Canchim**  
**C.Postal, 339 - São Carlos,SP - 13560-970**  
**Tel.: (16) 261 5611 - Fax: (16) 261 5754**  
**Email: [sac@cppse.embrapa.br](mailto:sac@cppse.embrapa.br)**  
**Home Page: [www.cppse.embrapa.br](http://www.cppse.embrapa.br)**

**Empresa Brasileira de Pesquisa Agropecuária - Embrapa  
Centro de Pesquisa de Pecuária do Sudeste - CPPSE  
Ministério da Agricultura, Pecuária e Abastecimento**

# **Anais**

## **IV Curso de Técnicas de Biologia Molecular Aplicadas à Produção Animal**

**Editado por:**

**Luciana Correia de Almeida Regitano  
Márcia Cristina de Sena Oliveira  
Luiz Lehmann Coutinho**

**São Carlos, SP**

Embrapa Pecuária Sudeste

Exemplares desta publicação podem ser solicitados à

Embrapa Pecuária Sudeste

Rod. Washington Luiz, km 234

Caixa Postal 339

CEP: 13560-970 São Carlos, SP

Telefone: (0xx16) 2615611

Fax: (0xx16) 2615754

"Home Page": <http://www.cppse.embrapa.br>

Endereço Eletrônico: [sac@cppse.embrapa.br](mailto:sac@cppse.embrapa.br)

Tiragem: 50 exemplares

---

Curso de Técnicas de Biologia Molecular Aplicadas à Produção  
Animal (4.: 2003 : São Carlos, SP)

Anais.../Editado por Luciana Correia de Almeida Regitano,  
Márcia Cristina Sena Oliveira, Luiz Lehmann Coutinho. – São  
Carlos: Embrapa Pecuária Sudeste, 2003.

1. Biologia molecular – Técnicas - Produção Animal. I.  
Oliveira, Marcia C. S. II. Coutinho, Luiz L. III. Título.

CDD 572.8 ( 21.ed.)

---

© Embrapa 2003

EMBRAPA/CPPSE/SÃO CARLOS
AIN
Valor Aquisição CRS
Data Aquisição 11/07/2003
Nº N. Fiscal Fatura
Fornecedor CPPSE
N. Ordem Compra
Origem Alacás
N. de Tombo 2003.03771

# ÍNDICE

<b>Aplicação de marcadores moleculares no melhoramento animal.....</b>	<b>01</b>
<i>Luciana Correia de Almeida Regitano</i>	
<b>Utilização da técnica de PCR em parasitologia veterinária .....</b>	<b>08</b>
<i>Márcia Cristina de Sena Oliveira</i>	
<b>Principais estratégias utilizadas no mapeamento do genoma de animais domésticos .....</b>	<b>18</b>
<i>Maria Elisabete J. Amaral</i>	
<b>Delineamentos experimentais para o mapeamento de QTL em animais .....</b>	<b>21</b>
<i>Mário Luiz Martinez</i> <i>Marcos Vinicius G. Barbosa da Silva</i>	
<b>Métodos e modelos estatísticos usados no mapeamento de QTLs .....</b>	<b>41</b>
<i>Marcos Vinicius G. Barbosa da Silva</i> <i>Mário Luiz Martinez</i>	

## Aplicação de marcadores moleculares no melhoramento animal

Luciana Correia de Almeida Regitano  
Embrapa Pecuária Sudeste

### Introdução

O melhoramento animal tem sido responsável por mudanças significativas na produção e na qualidade dos produtos de origem animal. Uma linhagem de frango de corte de 1991 chega a ser três vezes mais pesada do que um linhagem de 1957 sob a mesma dieta e à mesma idade (Haley, 1995). A taxa de crescimento da produção anual de leite por vaca tem aumentado nos últimos anos na América do Norte. Os valores de DEP de fêmeas nascidas em 1986 foram aproximadamente 135 kg superiores aos das fêmeas nascidas em 1985 (Wiggans, 1991).

Nos últimos 10 anos, com os desenvolvimentos metodológicos e científicos alcançados pelos projetos genoma humano e de espécies modelo, surge a possibilidade de se utilizar ferramentas moleculares para auxiliar o melhoramento animal. Aliadas às metodologias tradicionais, as novas técnicas deverão aumentar ainda mais o progresso genético que vem sendo observado em animais domésticos. O uso de marcadores moleculares, principalmente de DNA, permite que o potencial genético de um animal seja determinado com maior precisão e antes mesmo da expressão do seu fenótipo.

Projetos de mapeamento genômico em animais domésticos aliados ao mapeamento comparativo deverão fornecer as bases para a manipulação do genoma dos animais de acordo com as necessidades e exigências do mercado. Em bovinos, mapas de ligação compostos por cerca de 3.600 marcadores posicionados com média de intervalos inferior a 2,5 centimorgans (Bishop et al., 1994; Barendse et al., 1997; Kappes et al., 1997; <http://locus.jouy.inra.fr>) permitem o mapeamento de genes envolvidos na herança de caracteres quantitativos (QTLs).

Apesar do evidente progresso alcançado na tecnologia genômica e da identificação de várias regiões cromossômicas responsáveis pela herança de algumas características já ser uma realidade, a questão do mapeamento de QTLs é complexa em virtude das variações de fase de ligação marcador-QTL entre populações, da presença de epistasia e da magnitude dos efeitos, que precisam ser suficientes para justificar sua aplicação no melhoramento animal. A identificação e clonagem dos QTLs em si já pode ser vislumbrada por recentes resultados (Wilson et al., 2001; Blott et al., 2003), o que deve contribuir para solucionar parte desses problemas. Além disso, a aplicação de marcadores moleculares no diagnóstico de doenças hereditárias, na identificação individual e na verificação de parentesco vem deixando sua contribuição como ferramenta auxiliar nos programas de melhoramento.

## Tipos de marcadores moleculares

Marcadores genéticos são características de herança mendeliana simples que possibilitam a inferência do genótipo a partir do fenótipo do indivíduo, permitindo que a segregação do gene marcador seja acompanhada. A análise de segregação requer a existência de pelo menos duas formas alélicas, ou seja, a existência de polimorfismo no loco marcador. Inicialmente, as características disponíveis para esse tipo de análise eram as mutações que produziam alterações morfológicas, como o nanismo, a ausência de asas em *Drosophila* e a ausência de pêlos em camundongos. Entretanto, tais mutações são pouco freqüentes nas populações naturais, nas quais a maior parte da variação genética é de caráter contínuo (Tanksley, 1993). Além disso, mutações como as citadas acima, freqüentemente comprometem a adaptação do indivíduo, o que reduz sua utilidade em estudos de comparação entre populações.

Com o desenvolvimento de métodos de análise do DNA, a possibilidade de caracterizar as variações individuais na seqüência do material genético permitiu uma grande ampliação do leque de marcadores genéticos para incluir diferentes tipos de marcadores moleculares, classificados de acordo com a natureza da variação observada.

Uma das mais marcantes contribuições ao estudo de marcadores moleculares foi o desenvolvimento da técnica de reação em cadeia da polimerase (PCR). Essa técnica foi desenvolvida por Mullis, em 1983 (Mullis, 1990), mas sua importância só ficou demonstrada com a publicação dos primeiros trabalhos de aplicação da PCR em diagnóstico de doenças (Saiki *et al.*, 1985). A técnica consiste na replicação do DNA *in vitro*, catalisada por uma DNA polimerase termoestável, em presença dos quatro tipos de desoxinucleotídeos (dATP, dCTP, dTTP e dGTP) e de oligonucleotídeos sintéticos ou "primers", complementares às extremidades da região do DNA que se deseja amplificar. A técnica possibilita o estudo de quantidades ínfimas de DNA em um reduzido espaço de tempo, desde que se conheça parte da seqüência a ser analisada.

Os polimorfismos de comprimentos de fragmentos de restrição (RFLPs) são marcadores geralmente dialélicos, resultantes da variação do número de cortes efetuados por determinada enzima de restrição em função do número de sítios de restrição presentes ao longo da molécula de DNA. Essa variação é o resultado de mutações de ponto, que eliminam ou criam sítios de restrição para determinada enzima, podendo também resultar de eventos de inserção ou de deleção entre dois sítios de restrição adjacentes. Esses marcadores foram originalmente analisados pela técnica de "Southern blot" e sua identificação passava por uma longa etapa de desenvolvimento de sondas e teste destas contra um painel de DNA de diferentes indivíduos tratados com diversas endonucleases de restrição até a identificação de combinações polimórficas. A partir de sua caracterização, é possível transformar um marcador RFLP tradicional em um PCR-RFLP, bastando para isso desenvolver "primers" complementares à região que flanqueia o sítio polimórfico de restrição.

Uma classe de marcadores de grande importância em genética animal são os microssatélites ou SSR (simple sequence repeats). Esses marcadores caracterizam-se por repetições em *tandem* de um a cinco nucleotídeos, amplamente distribuídos pelo genoma, localizadas dentro de regiões de seqüência única. Cada bloco de repetições é geralmente menor do que 100 pares de nucleotídeos (Tautz, 1989). O intenso polimorfismo está relacionado a diferenças no número de repetições presentes em diferentes alelos, provavelmente resultantes de erros no deslocamento da DNA polimerase durante a replicação do DNA.

Os microssatélites podem ser amplificados de maneira específica pela técnica de PCR, utilizando *primers* que contém parte da seqüência flanqueadora. A amplificação resulta em produtos de diferentes tamanhos, em função do número de cópias da seqüência repetitiva delimitada pelos *primers*. A maior dificuldade técnica está relacionada à identificação dos genótipos, principalmente nos microssatélites constituídos por repetições de um a dois nucleotídeos, nos quais dois alelos adjacentes diferem entre si por apenas um ou dois pares de bases, respectivamente. Entretanto, sua elevada freqüência, ampla distribuição pelo genoma e alto conteúdo de polimorfismo fazem dos microssatélites marcadores ideais para mapeamento, identificação individual e verificação de relações de parentesco.

Recentemente, com a automação dos métodos de sequenciamento do DNA, uma nova categoria de marcadores vem ganhando destaque, os polimorfismos de um nucleotídeo ou SNP (single nucleotide polymorphism). O polimorfismo resulta de mutações de ponto que podem ser transições, transversões, inserções ou deleções. Sua detecção pode ser feita por sequenciamento, PCR alelo-específico, hibridação alelo-específica em chips de DNA ou espectrometria de massa. Alguns SNPs podem ser também convertidos em PCR-RFLP, quando a mutação de ponto altera um sítio de restrição (Vignal et al., 2002).

Apesar de geralmente dialélicos, os SNPs são marcadores atraentes por se encontrarem em praticamente qualquer região do genoma ou seqüência de interesse, como por exemplo, em genes candidatos à associação com uma característica. Outras características importantes são a maior estabilidade, quando comparados aos marcadores microssatélites.

Sua maior aplicação tem sido no mapeamento fino de QTLs, quando a informação de microssatélites na região que contém o QTL foi esgotada e os genes com possível papel fisiológico sobre a característica, contidos nessa região, foram identificados. Esses genes são analisados quanto à presença de SNPs que, uma vez encontrados, são avaliados quanto à associação com a característica fenotípica.

### **Diagnóstico de doenças hereditárias**

A aplicação de análise molecular para identificar mutações deletérias é de extrema utilidade, visto que muitas dessas mutações são recessivas e se encontram em baixa freqüência nas populações. A alternativa tradicional para a seleção de reprodutores livres de uma determinada mutação seria o teste de progênie com um número tanto maior de fêmeas

quanto menor a frequência do alelo mutante na população. Com o diagnóstico molecular, é possível determinar se o animal é ou não portador dessa mutação antes mesmo do início da reprodução ou ainda, no caso de transferência de embriões, na seleção dos embriões a serem transferidos.

Alguns exemplos de mutações para as quais já existem testes diagnósticos com base na análise molecular são a Deficiência de Adesão dos Leucócitos (BLAD), a citrulinemia e a deficiência de uridina monofosfato sintetase (DUMPS), as quais acometem bovinos da raça Holandesa. As informações sobre o estado de portador ou não dessas mutações é disponibilizada pela maioria das empresas que comercializam sêmen de animais dessa raça. Com as informações geradas pelos projetos genoma humano e de espécies modelo, um grande número de doenças hereditárias dos animais domésticos têm tido sua etiologia esclarecida, o que possibilitará o desenvolvimento de testes diagnósticos moleculares. As principais doenças hereditárias dos animais são descritas no endereço <http://morgan.angis.su.oz.au>.

### **Verificação de parentesco**

A correta documentação das relações de parentesco entre os animais é fator fundamental para o sucesso do melhoramento animal, uma vez que a premissa básica da hereditariedade é a transmissão vertical dos caracteres, ou seja, os alelos favoráveis só serão transmitidos à progênie verdadeira dos reprodutores selecionados. Além disso, muitos dos métodos de estimativa do valor genético de um animal utilizam as informações de produção de parentes.

O controle genealógico é também uma prática que permite a manutenção de raças como *pools* gênicos isolados, sem efeitos de introgressão de genes de outras raças.

Marcadores moleculares são reconhecidamente eficazes para a identificação individual e verificação de parentesco, sendo tanto mais eficazes quanto mais polimórficos e reprodutíveis tecnicamente. A possibilidade de utilização da técnica de PCR para sua análise é um ponto importante para casos de investigação forense por viabilizar a análise de marcadores a partir de poucas moléculas de DNA, que podem ser obtidas a partir de praticamente qualquer tecido do indivíduo.

### **Mapeamento de QTLs**

A existência de genes principais, que contribuem com grande parte da variação fenotípica de um caráter quantitativo, tem sido demonstrada. Alguns exemplos são o gene da hipertrofia muscular (*mh*) em bovinos, a mutação no receptor da Ryanodina dos suínos e o gene de fertilidade (*Fec*) dos ovinos. Essa situação, entretanto, pode ser considerada como exceção uma

vez que a maioria dos caracteres quantitativos deve ser controlada por um grande número de genes, cada um com pequeno efeito sobre o caráter (Massey & Georges, 1992).

Thoday (1961) formulou a hipótese de que se a segregação de um gene marcador pudesse ser utilizada para identificar e estimar o efeito de um poligene, e se um número suficiente de marcadores estivesse distribuído pelo genoma de uma espécie, seria possível mapear e caracterizar todos os poligenes que afetam um caráter quantitativo.

Para alcançar esse objetivo, duas estratégias vêm sendo utilizadas. A primeira utiliza marcadores aleatórios para a construção de mapas genéticos saturados. Esses marcadores são então avaliados em estudos de correlação com características de interesse econômico e utilizados no mapeamento QTLs. O sucesso dessa estratégia depende primariamente da construção de mapas de ligação detalhados, com marcadores dispostos a intervalos inferiores a 20 cM. Essa estratégia tem sido utilizada com sucesso em diversos estudos, como o de Ron et al. (1994) no qual um QTL com um efeito de 0,2 desvios-padrões fenotípicos para produção de leite e 0,16 desvios-padrões para proteína total foi mapeado no cromossomo 21 dos bovinos. Georges et al. (1995), encontraram fortes evidências da presença de QTLs para características de produção de leite nos cromossomos 1, 6, 9, 10 e 20 dos bovinos. Os efeitos dos diferentes QTLs contribuíram com 11 a 52% da variância total dentro de famílias de meio-irmãos.

A segunda estratégia baseia-se na avaliação de polimorfismos em genes sabidamente importantes para a característica que se pretende estudar. Os animais são classificados quanto às variações genéticas observadas nesses genes (RFLPs, SNPs, polimorfismos de conformação) e a média de produção das diferentes classes genotípicas é comparada.

O principal requisito desta estratégia é o conhecimento da fisiologia da característica fenotípica que se deseja estudar. A principal limitação é que, apesar de haver grande volume de sequências conhecidas na maioria dos genomas de espécies domésticas, o número de genes com função conhecida ainda é reduzido. Essa limitação tem sido sobrepujada pela análise comparativa com genomas mais conhecidos.

A associação das duas estratégias é freqüentemente necessária para se obter o mapeamento fino e clonagem posicional do QTL. Resultados de estudos integrando estratégias de mapeamento por intervalos, genes candidatos, mapeamento comparativo e técnicas de análise de expressão gênica têm demonstrado a possibilidade de clonagem posicional de um QTL. Um exemplo elegante é o estudo conduzido por Wilson et al. (2001) que atribuiu o fenótipo de alta taxa de ovulação de ovelhas da linhagem Booroola à uma mutação no gene do receptor da proteína de morfogênese óssea (BMPR-IB).

## **Considerações finais**

Os primeiros esforços no sentido de identificar genes que controlam características de interesse econômico em animais foram limitados pela falta de marcadores genéticos informativos. Com o desenvolvimento da genética molecular, bons resultados vêm sendo

obtidos, demonstrando a existência de QTLs para diferentes características nos diversos cromossomos dos bovinos.

Os conhecimentos acumulados na área da genética molecular devem-se em grande parte ao desenvolvimento científico decorrente dos projetos de mapeamento do genoma humano. A utilização dessas informações por meio de mapeamento comparativo e homologia de sequências têm contribuído, e deverá ser fundamental para futuro desenvolvimento dos projetos na área animal.

## Referências bibliográficas

Barendse, W., Vaiman, D., Kemp, S.J., Sugimoto, Y., Armitage, J.L., Williams, J.L., Sun, H.S., Eggen, A., Agaba, M., Aleyasin, S.A., Band, M., Bishop, M.D., Buitkamp, J., Byrne, K., Collins, F., Cooper, L., Copettiers, W., Denys, B., Drinkwater, R.D., Easterday, K., Elduque, C., Ennis, S., Erhardt, G., Ferretti, L., Flavin, N., Gao, Q., Georges, M., Gurung, R., Harzilius, B., Hawkins, G., Hetzel, J., Hirano, T., Hulme, D., Jorgensen, C., Kessler, M., Kirkpatrick, B.W., Konfortov, B., Kostia, S., Kuhn, C., Lenstra, J.A., Leveziel, H., Lewin, H.A., Leyhe, B., Lil, L., Burriel, I.M., Mcgraw, R.A., Miller, J.R., Moody, D.E., Moore, S.S., Nakane, S., Nijman, I.J., Olsaker, I., Pomp, D., Rando, A., Ron, M., Shalom, A., Teale, A.J., Thieven, U., Urquhart, B.G.D., Vage, D.I., Van De Weghe, A., Varvio, S., Velmala, R., Vilkki, J., Weikard, R., Woodside, C., Womack, J.E., Zanotti, M., Zaragoza, P. A medium-density genetic linkage map of the bovine genome. **Mammalian Genome**, **8**:21-28, 1997.

Bishop, M. D., Kappes, S. M., Keele, J. W., Stone, R. T., Sunden, S.L.F., Hawkins, G. A., Toldo, S. S., Fries, R., Grosz, M. D., Yoo, J., Beattie, C. W. A genetic linkage map for cattle. **Genetics**, **136**: 619-639, 1994.

Blott, S.; Kim, J.J.; Moisiu, S.; Schmidt-Küntzel, A; Cornet, A.; Berzi, P.; cambisano, N.; Ford, C.; Grisart, B.; Johnson, D.; Karim, L.; Simon, P.; Snell, R.; Spelman, R.; Wong, J.; Vilkki, J.; Georges, M.; farnir, F.; Coppieters, W. Molecular dissection of a quantitative trait locus: a phenylalanine-to-tyrosine substitution in the transmembrane domain of the bovine growth hormone receptor is associated with major effect on milk yield and composition. **Genetics**, **163**: 253-266.

Georges, M.; Nielsen, D.; Mackinnon, M.; Mishra, A.; Okimoto, R.; Pasquino, A.T.; Sargeant, L.S.; Sorensen, A.; Steele, M.R.; Zhao, X.; Womack, J.E.; Hoeschele, I.. Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing. **Genetics**, **139**: 907-920, 1995.

Halley, C. S. Livestock QTLs – bringing home the bacon? **Trends in Genetics**, **11**: 488 – 492, 1995.

<http://locus.jouy.inra.fr> - INRA Biotechnology Laboratories Home Page. Consultado em maio de 2003.

- Kappes S.M.; Keele J.W.; Stone R.T.; McGraw R.A.; Sonstegard T.S.; Smith T.P.; Lopez-Corrales N.L.; Beattie, C.W. A second-generation linkage map of the bovine genome. **Genome Research**, 7(3):235-49, 1997.
- Massey, J.M. & Georges, M. Genmark's approach to marker-assisted selection. **Animal Biotechnology**, 3: 95-109, 1992.
- Mullis, K. B.. The unusual origin of the polymerase chain reaction. **Scientific American**, 262: 36 - 42, 1990.
- Ron, M.; Band, M.; Yanai, A.; Weller, J. I. Mapping quantitative trait loci with DNA microsatellites in a commercial dairy cattle population. **Animal Genetics**, 25: 259-264, 1994.
- Saiki, R. K.; Scharf, S.; Faloona, F.; Mullis, K. B.; Horn, G. T.; Erlich, H. A.; Arnheim, N. Enzymatic amplification of  $\beta$ -globin genomic sequences and restriction site analysis for diagnosis of Sickle Cell Anemia. **Science**, 230: 1350-1354, 1985.
- Tanksley, S. D. Mapping polygenes. **Annual Review of Genetics**, 27: 205-233, 1993.
- Tautz, D. Hypervariability of simple sequences as a general source for polymorphic DNA markers. **Nucleic Acids Research**, 17: 6463-6471, 1989.
- Thoday, J.M. Location of polygenes. **Nature**, 191: 368-370, 1961.
- Vignal, A.; Milan, D.; SanCristobal, M.; Eggen, A. A review on SNP and other types of molecular markers and their use in animal genetics. **Genet Sel Evol.**, 34: 275-305, 2002.
- Wiggans, G.R. National genetic improvement programs for dairy cattle in the United States. **Journal of Animal Science**, 69 (9):3853-60, 1991.
- Wilson, T.; Wu, X.Y.; Juengel, J.L.; Ross, I.K.; Lumsden, J.M.; Lord, E.A.; Dodds, K.G.; Walling, G.A.; McEwan, J.C.; O'Connell, A.R.; McNatty, K.P.; Montgomery, G.W. Highly prolific Booroola sheep have a mutation in the intracellular kinase domain of bone morphogenetic protein IB receptor (ALK-6) that is expressed in both oocytes and granulosa cells. **Biology of reproduction**, 64: 1225-1235, 2001.

## Utilização da técnica de PCR em parasitologia veterinária

Márcia Cristina de Sena Oliveira  
Embrapa Pecuária Sudeste

### Introdução

Com o avanço rápido dos estudos de genética molecular muita ênfase tem sido dada a predisposição genética a doenças. Resultados de pesquisa identificam locus associados a predisposição a várias enfermidades, e de certa maneira, o estudo das doenças parasitárias e infecciosas ficaram em segundo plano. No entanto, essas doenças podem ser vistas como "desordens genéticas adquiridas horizontalmente", no qual seqüências de ácidos nucleicos originárias de agentes patogênicos são responsáveis por produzir reação inflamatória e doença (Relman & Persing, 1996).

As técnicas microscópicas permanecem como padrão para o diagnóstico de muitos parasitas, apesar de serem trabalhosas e necessitarem de pessoal bem treinado para a identificação e interpretação dos exames, particularmente para parasitas cuja morfologia é similar em diferentes espécies ou estão em baixo número na amostra.

Os métodos de diagnóstico baseados no genótipo apresentam vantagens em relação aos que utilizam características fenotípicas. As técnicas de diagnóstico que detectam o DNA do parasita reduzem a subjetividade inerentes a interpretação de dados morfológicos e biológicos e o DNA essencialmente não varia através de estágios diferentes do ciclo de vida dos parasitas causadores de doenças.

Na prática, seqüências alvo de DNA ou RNA específicos de parasitas são detectados basicamente por dois métodos: hibridização da seqüência homóloga com DNA marcado e amplificação enzimática da seqüência alvo por meio da Reação em Cadeia da Polimerase (PCR). Essas técnicas têm possibilitado o avanço significativo no conhecimento relativo a sistemática, epidemiologia, imunologia, interação parasita-hospedeiro e desenvolvimento de vacinas e medicamentos, entre outros aspectos de interesse.

### Genoma de parasitas de interesse veterinário

O seqüenciamento do genoma de muitos microrganismos e parasitas serve de base para o estudo não somente da espécie analisada, como de muitas espécies relacionadas a elas. O *Caenorhabditis elegans* foi o primeiro organismo multicelular a ter seu genoma seqüenciado e representou uma imensa fonte de pesquisa para muitos outros helmintos (Anon, 1998). Todo o conhecimento gerado a partir do seqüenciamento do genoma de parasitas serve como base para análises funcionais de novos genes descobertos. Os

resultados obtidos no projeto *schistosoma* disponibilizaram informações que foram utilizadas no estudo de parasitas de interesse veterinário, como *Fasciola hepatica*. Outros projetos genoma também direcionados para doenças humanas, tiveram ampla aplicação em pesquisas de parasitas de animais, como os projetos *Leishmania*, *Plasmodium* e *Entamoeba* e *Brugia malayi*. Atualmente, cerca de oito projetos genoma de parasitas estão em execução e muitos outros fazem parte de uma lista de prioridades como *Eimeria* spp. e *Toxoplasma goondi* (Prichard & Tait, 2001).

### **Desenvolvimento de “primers” para reações de PCR**

Algumas técnicas podem ser utilizadas para o desenvolvimento de “primers” ou seqüências iniciadoras para PCR. A seqüência do gene 18S rDNA é comumente utilizada para o desenvolvimento de “primers”, devido a sua disponibilidade (<http://rrna.uia.ace.be>) para um grande número de protozoários e outros parasitas (Morgan & Thompson, 1998). Os “primers” podem ser facilmente designados a partir dessas seqüências, porém, devido a grande conservação observada nesses genes a ocorrência de reações cruzadas com outros organismos pode ocorrer. Um método alternativo é por meio da construção de bibliotecas de DNA genômico. Esse método é caro e demanda tempo, necessitando de grande quantidade de DNA, que pode ser difícil de se obter dependendo do microrganismo em questão. O uso da técnica de amplificação de DNA polimórfico ao acaso (RAPD) desenvolvida independentemente por Williams et al. (1990) e Welsh & McClelland (1990) tem sido uma opção mais simples para ser utilizada. Essa técnica detecta polimorfismos em seqüência de nucleotídeos nos testes de PCR, sem a necessidade de qualquer informação prévia. Como essa técnica é baseada em PCR, pequenas quantidades de DNA são suficientes para a análise. Muitos dos produtos gerados pelo RAPD-PCR são derivados de seqüências repetitivas do DNA, portanto espécie-específicas e dessa maneira adequados para o delineamento de técnicas de diagnóstico (Williams et al., 1990). Dessa maneira bandas geradas por RAPD-PCR podem ser eluídas do gel, sua especificidade pode ser testada por hibridização e então clonada e seqüenciada e a partir dessas seqüências os “primers” podem ser designados e sintetizados (Morgan & Thompson, 1998).

### **Preparação das amostras**

O isolamento e purificação de ácidos nucleicos de amostras de agentes de doenças é uma importante etapa para se conseguir alta eficiência de amplificação e especificidade nas reações de PCR. Para as reações de hibridização a mesma observação se aplica, já que pode ser difícil de se obter amostra de DNA genômico em quantidade suficiente e pura, de muitos estágios de alguns parasitas. Os helmintos são particularmente difíceis de se obter boa

amostra de DNA, devido a presença de cutícula e substâncias que precipitam junto com os ácidos nucleicos durante o seu isolamento (Gasser et al., 1998).

Existem vários métodos para extração de ácidos nucleicos que podem ser usados para parasitas e microrganismos, no entanto, devido a pequena quantidade de DNA presente nesses agentes, é necessário a otimização do processo. Segundo Gasser et al. (1998) 0,1 pg de rDNA de *Oesophagostomum* spp. é uma quantidade adequada para ser utilizada em amplificação por meio de PCR.

Para extração de DNA, amostras contendo o agente são homogeneizadas em tampão contendo RNase, as células são submetidas a lise e a amostra é digerida com proteinase K. A partir daí o DNA pode ser extraído com fenol, precipitado com etanol e dissolvido em água purificada esterilizada ou tampão aquoso.

Precauções são necessárias na extração de RNA devido à sua característica de ser facilmente degradável. Dessa maneira o congelamento rápido em nitrogênio líquido é usado para inativar rapidamente as ribonucleases, que geralmente podem se manter estáveis por longos períodos, não requerendo co-fatores para sua ativação. Todo o material utilizado na colheita das amostras deve ser livre dessas enzimas, sendo que a utilização de luvas é obrigatória. Enquanto o material está ainda congelado um agente desnaturante como fenol ou guanidina deve ser adicionado, e dessa maneira, após a lise das células e digestão das proteínas, o RNA pode ser fracionado e separado de outras macromoléculas

As amostras de DNA ou RNA também podem ser obtidas por meio da utilização de kits para purificação, produzidos por diversos laboratórios. Estes kits podem tornar o processo de extração mais fácil e rápido.

### Reações de hibridização de ácidos nucleicos

O aquecimento de uma solução de DNA a temperaturas ao redor de 90-100° C assim como a exposição a pH extremamente alcalino, provoca o rompimento da estrutura de dupla hélice do ácido nucleico. Esse processo chamado de desnaturação foi considerado irreversível por um longo tempo, até que foi demonstrado que as fitas simples complementares podiam ser novamente dispostas em dupla hélice pelo processo reverso, denominado renaturação ou hibridização. A partir desse conhecimento fitas simples de DNA puderam ser utilizadas como sondas para se determinar a presença de qualquer seqüência de DNA ou RNA. Fragmentos de ácidos nucleicos podem ser identificados por transferência de RNA total ou de DNA genômico para membranas de nitrocelulose, nas quais as reações de hibridização são realizadas empregando-se sondas moleculares específicas ao gene de interesse. Tal procedimento experimental pode ser realizado com DNA e RNA, sendo denominados "southern" e "northern blotting", respectivamente (Gabriel, 2001).

A técnica de hibridização de DNA foi utilizada por Buening et al. (1990) para detectar DNA específico de *Babesia bigemina*. Esses autores construíram uma biblioteca de DNA

recombinante que continha seqüências repetitivas do genoma desse protozoário. Após a clonagem, a sonda selecionada hibridizou apenas com o DNA genômico de *B. bigemina*, não o fazendo com DNA de *B. bovis* e outros hemoparasitas. Além disso, verificaram que essas seqüências se conservavam em várias cepas geográficas como as provenientes do México, Porto Rico, Costa Rica e Kênia. Nos ensaios sobre sensibilidade verificou-se que parasitemias da ordem de 0,001% podiam ser detectadas com o uso dessa sonda. A técnica de hibridização também foi utilizada por Hodgson et al. (1992) para estimar a susceptibilidade de ninfas e adultos de *Boophilus microplus* à infecção por *B. bigemina*. Nessas técnicas, como o DNA hibridizado é detectado por marcação radioativa, alguns inconvenientes como os relacionados ao manuseio de isótopos radioativos, aos altos custos do equipamento e à exigência de pessoal especializado, constituem graves limitações. Esses problemas incentivaram o desenvolvimento de testes utilizando marcação não radioativa com digoxigenina. Entretanto, esta metodologia quando aplicada ao diagnóstico de hemoparasitas apresentou o inconveniente de reduzir a sensibilidade do teste (Aboytes et al., 1991). Os testes de hibridização de DNA para fins de diagnóstico foram abandonados porque dependem essencialmente da quantidade de DNA do agente presente na amostra e a sensibilidade desses testes são comparáveis à dos exames microscópicos.

### Reação em cadeia da polimerase (PCR)

A técnica de Reação em Cadeia da Polimerase (PCR), desenvolvida por Mullis & Faloona (1987), na qual uma determinada região do genoma de qualquer organismo pode ser multiplicada em milhões de cópias, possibilitou o desenvolvimento de técnicas de diagnóstico muito mais sensíveis e com alta especificidade. Essa técnica é considerada a base do desenvolvimento de uma nova geração de métodos de diagnóstico. A alta sensibilidade, especificidade, facilidade de execução e análise de um grande número de amostras simultaneamente, fazem dessa técnica uma opção atrativa para diagnóstico e utilização em levantamentos epidemiológicos, entre outros estudos com agentes de doenças (Morgan & Thompson, 1998).

As técnicas baseadas na reação de cadeia da polimerase (PCR) têm sido empregadas com sucesso em diversos estudos, como para a detecção de *Cryptosporidium* em água (States et al., 1997) e em casos clínicos de criptosporidiose (Leng et al., 1996); diagnóstico específico de *Plasmodium* spp. (Ayyanathan & Datta, 1996) e estudos epidemiológicos em malária (Roper et al., 1996); diagnóstico diferencial entre *Entamoeba histolytica* e *Entamoeba dispar* (Troll et al., 1997); diagnóstico de *Tritrichomonas foetus* (Ho et al., 1994); diferenciação de *Toxoplasma gondii* e *Neospora caninum* (Lally et al., 1996); detecção de *Trypanosoma* spp. no hospedeiro vertebrado (Katakura et al., 1997) e sua prevalência nos hospedeiros invertebrados (Masiga et al., 1996), entre outros.

Testes baseados em PCR, desenvolvidos para a detecção de *Babesia* spp., têm demonstrado sensibilidade 100 a 1.000 vezes maior que o limiar de detecção em microscopia

óptica (Böse et al., 1995). Com isso, oligonucleotídeos são disponíveis para a detecção de diferentes espécies como *Babesia gibsoni* (Fukumoto et al., 2001), *Babesia microti* (Skotarczak & Cichocka, 2001), *Babesia equi* e *Babesia caballi* (Battsetseg et al., 2001, 2002) e têm sido utilizados para pesquisa dos protozoários tanto nos hospedeiros vertebrados como nos invertebrados.

Figuroa et al. (1992) desenvolveram uma técnica de PCR específica para *B. bigemina*, utilizando um fragmento clonado de rDNA repetitivo para designar dois pares de "primers". O primeiro par, foi utilizado para amplificar um fragmento de DNA com 278 pares de bases (pb) e o segundo, para hibridizar com um sítio interno da seqüência inicialmente amplificada. A reação de amplificação da seqüência do DNA de *B. bigemina* obtida de cultura "in vitro", seguido da reação de hibridização do ácido nucleico amplificado, mostrou que o fragmento de 278 pares de bases pôde ser detectado quando quantidades tão pequenas quanto 100 fg de DNA genômico foram utilizados. Padrões idênticos de amplificação foram obtidos em testes em que se utilizaram várias cepas de *B. bigemina*, de diferentes procedências. Além disso, como essa seqüência não amplificou o DNA de amostras de *B. bovis*, *Anaplasma marginale*, seis espécies de bactérias e de leucócitos bovinos, mostrou-se bastante específica.

A técnica de PCR desenvolvida por Figuroa et al. (1992) foi utilizada em vários estudos conduzidos em várias partes do mundo: Hermans et al. (1994); Smeenk et al. (2000) no Zimbábue e Almeria et al. (2001) na Espanha.

Recentemente, Oliveira (2002) utilizou os "primers" designados por Figuroa et al., (1993) para desenvolver um teste de "nested PCR" (nPCR) para o diagnóstico de *Babesia bovis* e *Babesia bigemina* em amostras de sangue de bovinos, em fêmeas adultas e ovos de *Boophilus microplus*. Os "primers" internos de hibridização, foram utilizados para uma nova reação de PCR. A sensibilidade estimada para as reações de PCR para *Babesia bigemina* foi de  $6 \times 10^3$  eritrócitos parasitados em  $18 \times 10^6$  eritrócitos, que corresponde a uma parasitemia de 0,00003%. Nas reações de nPCR a sensibilidade estimada foi de 0,0000003%.

### PCR "multiplex"

A técnica de "multiplex" foi desenvolvida com a finalidade de, com um único teste, altamente específico, promover a diferenciação entre várias espécies ou gêneros simultaneamente (Zarlenga & Higgins, 2001). Essa forma de PCR envolve a amplificação simultânea de mais de uma seqüência alvo por reação, pela mistura de múltiplos pares de "primers". Essa técnica é especialmente útil para análise de amostras que contenham parasitas cuja distinção morfológica é difícil ou se apresentem em número muito baixo. Desse modo, "primers" altamente específicos são usados para amplificar seqüências conhecidas do DNA, produzindo padrões únicos para cada espécie.

Vários trabalhos foram conduzidos utilizando essa técnica para diferenciar parasitas: *Taenia saginata* de *Taenia asiatica* (Zarlenga et al., 1991), *Haemonchus contortus* e

*Haemonchus placei* (Zarlenga et al., 1994), espécies do gênero *Trichinella* (Zarlenga et al., 1999), espécies do gênero *Leishmania* (Belli et al., 1998).

Figuroa et al. (1993) desenvolveram um PCR "multiplex" para detecção de *B. bovis*, *B. bigemina* e *Anaplasma marginale* em sangue bovino. A sensibilidade do teste, avaliada por meio da hibridização do ácido nucleico, foi de 0,00001% de eritrócitos infectados para *B. bovis* e *B. bigemina* e 0,0001% para *Anaplasma marginale*.

## RAPD

A análise de DNA polimórfico amplificado ao acaso (RAPD) é usada para comparar diferenças e similaridades no DNA amplificado em reações de PCR, usando "primers" selecionados ao acaso. Os fragmentos sintetizados são separados e visualizados por eletroforese, de modo que a natureza polimórfica dos produtos amplificados de diferentes amostras podem ser comparadas (Prichard, 1997). A análise de RAPD tem sido utilizada em estudos de caracterização de espécies ou isolados de *Trypanosoma* de ruminantes e suínos (Dirie et al., 1993; Waimtumbi & Murphy, 1993).

## RT-PCR

A técnica de RT-PCR envolve a síntese de um transcrito de DNA complementar (cDNA) a partir de moléculas de RNA, por meio da utilização da enzima Transcriptase Reversa. O DNA sintetizado é então submetido a amplificação por meio da técnica de PCR, utilizando "primers" específicos que se ligam especificamente com as moléculas de mRNA do gene de interesse (Alvares, 2001). A utilização do RNA contido na amostra, tem como principal vantagem o fato de aumentar a sensibilidade do teste, já que esse ácido nucleico está presente em maior quantidade (cerca de 50%) em uma célula típica, quando comparado ao DNA (Zarlenga et al., 2001).

As técnicas de RT-PCR quantitativas e semi-quantitativas são consideradas de grande importância para a análise da expressão gênica em diferentes estágios do desenvolvimento de parasitas. Estudos recentes têm aplicado essa técnica para caracterizar a expressão de genes específicos de estágios de parasitas cestodos, que são importantes na predição da imunogenicidade de antígenos protetores (Waterkeyn et al., 1997).

## PCR quantitativo

A análise quantitativa por meio da cinética da técnica de PCR pode ser feita adicionando-se um corante fluorescente (SYBR Green™ ou brometo de etídeo) na reação de PCR. Com isso, conforme a reação progride, a amplificação produz quantidades crescentes de DNA dupla fita que se liga ao corante resultando em aumento da fluorescência. Plotando-se o aumento da fluorescência pelo número de ciclos, são produzidos gráficos de amplificação que

fornecem um panorama completo da PCR (Alvares, 2001). A técnica de PCR quantitativo utilizando o SYBR Green™ foi usada para quantificar *Borrelia burgdorferi* em modelos com animais de laboratório para a doença de Lyme (Morrison et al., 1999) e melhorar os métodos de diagnóstico para ehrlichiose (Edelman & Dumler, 1996).

Técnicas mais recentes como a dos "beacons" moleculares e DNA "arrays" deverão ser usadas no futuro com fins de diagnóstico e outros propósitos em parasitologia veterinária.

## Conclusões

As técnicas de amplificação de DNA encontram um amplo campo de aplicações em estudos de doenças infecciosas e parasitárias de interesse veterinário. Estudos recentes demonstram sua aplicação para identificação de agentes de doenças e sua caracterização genética, diagnóstico de infecções e infestações, estudos epidemiológicos, detecção de resistência a drogas, entre outros.

## Referências bibliográficas:

- Aboytes R., Buening, G.M., Figueroa, J.V., Vega, C.A., 1991. El uso de sondas de ADN para el diagnóstico de hemoparásitos. Rev. Cub. Cienc. Vet., .22 (3), 173-81.
- Almeria, S., CastellÀ, J., Ferrer, D., Ortuño, A., Estrada-Peña, A., Gutierrez, J.F., 2001. Bovine piroplasms in Minorca (Balearic Islands, Spain): a comparison of PCR-based and light microscopy detection. Vet. Parasitol., 99, 249-59.
- Alvares, L.E. 2001. Aplicação da RT-PCR nos Estudos de Expressão Gênica. IN: Regitano, L.C.A., Coutinho, L.L., eds. Biologia Molecular Aplicada à Produção Animal, Brasília, Embrapa, 2001, 135-150.
- Anon, 1998. *Caenorhabditis elegans* Sequencing Consortium. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. Science., 282, 2012-2018.
- Ayyanathan, K. & Datta, S., 1996. Identification and characterization of a generic DNA probes capable of detecting plasmodial infections in blood. Mol. Cel. Probes, 10, 273-8.
- Battsetseg, B., Xuan, X., Ikadai H., Jose, R.B.L., Byambaa, B., Bolbaatar, D., Battur, B., Battsetseg, G., Batsukh, Z., Igarashi, I., Nagasawa, H., Mikami, T., Fujisaki, K., 2001. Detection of *Babesia caballi* and *Babesia equi* in *Dermacentor nuttalli* adult tick. Int. J. Parasitol. 31, 384-386.
- Battsetseg, B., Lucero, S., Xuan, X., Claveria, F.G., Inoue, N., Alhassan, A., Kanno, T., Igarashi, I., Nagasawa, H., Mikami, T., Fujisaki, K., 2002. Detection of natural infection of *Boophilus microplus* with *Babesia equi* and *Babesia caballi* in Brazilian horses using nested-polymerase chain reaction. Vet. Parasitol. 107, 351-357.

- Belli, A., Rodriguez, B., Aviles, H., Harris, E., 1998. Simplified polymerase chain reaction detection of new world *Leishmania* in clinical specimens of cutaneous leishmaniasis. *Am. Trop. Med. Hyg.* 58, 102-109.
- Böse, R., Jorgensen, W.K., Dalgliesh, R.J., Friedhoff, K.T., DeVos, A.J., 1995. Current state and future trends in the diagnosis of babesiosis. *Vet. Parasitol.* 57, 61-74.
- Buening, G.M., Barbet, A., Myler, P., Mahan, S., McGuire, T.C., 1990. Characterization of a repetitive DNA probe for *Babesia bigemina*. *Vet. Parasitol.*, 36, 11-20.
- Dirie, M.F., Murphy, N.B., Gardiner, P.R., 1993. DNA Fingerprinting of *Trypanosoma vivax* isolates rapidly identifies intraspecific relationships. *J. Eukaryotic Microbiol.* 40, 132-134.
- Edelman, D.C., Dumler, J.S., 1996. Evaluation of an improved PCR diagnostic assay for human granulocytic ehrlichiosis. *Mol. Diagn.* 1, 41-49.
- Figuroa, J.V., Chieves, L.P., Johnson, G.S., Buening, G.M., 1992. Detection of *Babesia bigemina*-infected carriers by polymerase chain reaction amplification. *J. Clin. Microbiol.* 30 (10), 2576-2582.
- Figuroa, J.V., Chieves, L.P., Johnson, G.S., Buening, G.M., 1993. Multiplex polymerase chain reaction based assay for the detection of *Babesia bigemina*, *Babesia bovis* and *Anaplasma marginale* DNA in bovine blood. *Vet. Parasitol.*, 50, 69-81.
- Fukumoto, S., Xuan, X., Shigeno, S., Kimbita, E., Igarashi, I., Nagasawa, H., Fujisaki, K., Mikami, T., 2001. Development of a polymerase chain reaction method for diagnosing *Babesia gibsoni* infections in dogs. *J. Vet. Med. Sci.* 63 (9), 977-981.
- Gabriel, J.E., 2001. Emprego das técnicas de Biologia Molecular nos Estudos de Expressão Gênica. IN: Regitano, L.C.A., Coutinho, L.L., eds. *Biologia Molecular Aplicada à Produção Animal*, Brasília, Embrapa, 2001, 124-134.
- Gasser, R.B., Zhu, X.Q., Monti, J.R., Dou, L., Cai, X., Pozio, E., 1998. PCR-SSCP of de rDNA for the identification of *Trichinella* isolates from mainland China. *Mol. Cell. Probes.* 12, 27-34.
- Hermans, P., Dwinger, R.H., Buening, G. M., Herrero, M.V., 1994. Seasonal incidence and hemoparasite infection rates of Ixodid ticks (Acari: Ixodidae) detached from cattle in Costa Rica. *Rev. Biol. Trop.* 42 (3), 623-632.
- Ho, M.S., Conrad, P.A., Lefebvre, R.B., Perez, E., Bondurant, R.H., 1994. Detection of bovine trichomoniasis with specific DNA probe and PCR amplification system. *J. clin. Microbiol.* 32, 98-104.
- Hodgson, J.L., Stiller, D., Douglas, P.J., Buening, G.M., Wagner, G.G., McGuire, C., 1992. *Babesia bigemina*: Quantification of infection in nymphal and adult *Boophilus microplus* using DNA probe. *Ex. Parasitol.* 74, 117-26.
- Katakura, K., Lubinga, C., Chitambo, H., Tada, Y., 1997. Detection of *Trypanosoma congolense* and *T. brucei* subspecies in cattle in Zambia by polymerase chain reaction from blood collected on a filter paper. *Parasitol. Res.* 83, 241-5.

- Lally, N.C., Jenkins, M.C., Dubey, J.P., 1996. Development a polymerase chain reaction assay for the diagnostics of neosporosis using the *Neospora caninum* 14-3-3 gene. *Mol. Bioch. Parasitol.*, 75, 169-78.
- Leng, X., Mosier, D.A., Oberst, R.D., 1996. Simplified method for recovery and PCR detection of *Cryptosporidium* DNA from bovine feces. *App. Environ. microbiol.* 62, 643-7.
- Masiga, D.K., McNamara, J.J., Gibson, W.C., 1996. A repetitive DNA sequence specific for *Trypanosoma (Nannomonas) godfreyi*. *Vet. Parasitol.* 62, 27-33.
- Morgan, U.M., Thompson, R.C.A., 1998. Molecular detection of parasitic protozoa. *Parasitology*, 117, 73-85.
- Morrison, T.B., Ma, Y., Weis, J.H., Weis, J.J., 1999. Rapid and sensitive quantification of *Borrelia burgdorferi* infected mouse tissues y continuous fluorescent monitoring of PCR. *J. Clin. Microbiol.* 37, 987-992.
- Mullis, K., Faloona, F., 1987. Specific synthesis of DNA in vitro via polymerase catalysed chain reaction. *Methods Enzymol.* 55, 335-350.
- Oliveira, M.C.S. **Avaliação da taxa de infecção por *Babesia bigemina* em bovinos mestiços leiteiros e em fêmeas e ovos de *Boophilus microplus* provenientes de área endêmica do Estado de São Paulo.** Botucatu, 2002. 96 p. Dissertação (Doutoramento). Faculdade de Medicina Veterinária e Zootecnia, Universidade Estadual Paulista, Botucatu.
- Prichard, R., 1997. Application of a molecular biology in veterinary parasitology. *Vet. Parasitol.* 71, 155-175.
- Prichard, R., Tait, A., 2001. The role of molecular biology in veterinary parasitology. *Vet. Parasitol.* 98, 169-194.
- Relman, D.A., Persing, D.H., 1996. Genotypic Methods for Microbial identification In: Persing, D. H., ed. *PCR Protocols for Emerging Infectious Diseases*, Washington, ASM Press, 1996, 4-31.
- Roper, C., Elhassan, I.M., Hviid, L., Giha, H., Richardson, W., Babiker, H., Satti, G.M.H., Theander, T.G., Aarnot, D.E., 1996. Detection of very low level *Plasmodium falciparum* infections using the nested polymerase chain reaction and a reassessment of the epidemiology of unstable malaria in Sudan. *Am. J. Trop. Med. Hyg.*, 54, 325-31.
- Skotarczak, B., Cichocka, A., 2001. The occurrence DNA of *Babesia microti* in ticks *Ixodis ricinus*. *Folia Biol.* 49 (3-4), 247-50.
- Smeenk, I., Kelly, P.J., Wray, K., Musuka, G., Trees, A.J., Jongejan, F., 2000. *Babesia bovis* and *Babesia bigemina* DNA detected in cattle and ticks from Zimbabwe by polymerase chain reaction. *J. S. Afr. Vet. Assoc.* 71 (1), 1-4.
- States, S., Stadterman, K., Ammon, L., Vogel, P., Baldizar, J., Wright, D., Conley, L., Sykora, 1997. J. Protozoa in river water: sources occurrence, and treatment. *J. Am. Water Works Assoc.* 89, 74-83.

- Troll, H., Marti, H., Weiss, N., 1997. Simple differential detection of *Entamoeba histolytica* and *Entamoeba dispar* in fresh stool specimens by sodium acetate-acetic acid-formalin concentration and PCR. J. Clin. Microbiol. 35, 1701-5.
- Waimtumbi, J.N., Murphy, N.B., 1993. Inter and intra-species differentiation of trypanosomes by genomic fingerprinting with arbitrary primers Mol. Biochem. Parasitol. 58, 181-186.
- Waterkeyn, J., Gauci, C., Cowman, A., Lighthowlers, M.W., 1997. Sequence analysis of a gene family encoding *Taenia ovis* vaccine antigens expressed during embryogenesis of eggs. Mol. Biochem. Parasitol. 86, 75-84.
- Welsh, J., McClelland, M., 1990. Fingerprinting genomes using PCR with arbitrary primers. Nucl. Acids Res. 18, 7213-7218.
- Williams, J.G.K., Kubelik, A.R., Livak, K.J., Rafalski, J.A., Tingey, S.V., 1990. DNA polymorphisms amplified by arbitrary are useful as genetic markers. Nucl. Acids Res. 18, 6531-6535.
- Zarlenga, D.S., Chute, M.B. Martin, A., Kapel, C.M.O., 1999. A multiplex PCR for unequivocal differentiation of six encapsulated and three non encapsulated genotypes of *Trichinella*. Int. J. Parasitol. 29, 141-149.
- Zarlenga, D.S., Higgins, J., 2001. PCR as a diagnostic and quantitative technique in veterinary parasitology. Vet. Parasitol. 101, 215-230.
- Zarlenga, D.S., MacManus, D.P., Fan, P.C., Cross, J.H., 1991. Characterization and detection of a newly described Asian Taeniid using closed ribosomal DNA fragments and sequence amplification by the polymerase chain reaction. Exp. Parasitol. 72,, 174-183.
- Zarlenga, D.S., Stringfellow, F., Nobary, M., Lichtenfels, J.R., 1994. Cloning and characterization of ribosomal RNA genes from three species of *Haemonchus* (Nematoda, Trichostrongyloidea) and identification of PCR primers for rapid differentiation. Exp. Parasitol. 78, 28-36.

## Principais estratégias utilizadas no mapeamento do genoma de animais domésticos

Dra. M. Elisabete J. Amaral  
IBILCE - UNESP, São José do Rio Preto

### Mapeamento do genoma

Objetivo do mapeamento genômico: é a indicação de um gene a um locus (região específica) no cromossomo. Pelo mapeamento as posições relativas dos genes e outros marcadores torna possível a construção de um mapa do cromossomo ou um mapa de todo o genoma. O mapeamento genômico tem sido usado para ajudar a entender e explorar a herança das características biológicas, particularmente associando a herança de uma característica com outra ou com um marcador genético adequado e pela correlação de diferenças no fenótipo com diferenças na estrutura do cromossomo.

A disciplina de mapeamento, análise e sequenciamento de genomas é denominada: "Genômica"

### Biotechnology aplicada aos animais domésticos:

A biotecnologia vêm sendo aplicada aos animais domésticos em duas grandes áreas:

- a) Genômica: identificação de genes relacionados com características, que pode ser subdividida em genômica estrutural e funcional.
- b) Transgênese: modificação de genes e inserção dentro de organismos hospedeiros (manipulação de genes/características).

A pesquisa genômica em animais domésticos está se desenvolvendo em 4 fases:

- 1- Construção de mapas com marcadores altamente informativos e genes;
- 2- uso destes mapas para fazer um scan em genomas de populações cruzadas segregantes para características comercialmente importantes, visando localizar QTL dentro de segmentos cromossômicos de 20-40 cM.
- 3- Identificação dos genes responsáveis por características, dentro destas regiões, através do uso de ESTs e bibliotecas de cDNA do tecido alvo.
- 4- Preencher o *gap* entre gene → fenótipo final

Tipos de mapas Genômicos:

Pode-se considerar 2 tipos diferentes de mapas:

- Mapas genéticos (ligação)
- Mapas físicos (moleculares)

Mapas Genéticos ou de ligação (linkage):

- Indicam genes ou marcadores à grupos de ligação;
- São organizados (calibrados) em unidades arbitrárias, refletindo a probabilidade da separação dos marcadores pela recombinação (meiose) ou pela fragmentação do genoma (híbridos irradiados).
- O princípio do mapeamento genético: a chance de dois loci aparentemente juntos no mesmo fragmento de DNA diminui com o aumento da distância entre eles.
- Mapa físico ou molecular: é criado pela ordenação de fragmentos clonados de DNA genômico, sendo organizados em unidades reais (bp, Kb, Mb) – FISH, clone contigs, etc.

***Genética de células somáticas:***

Características importantes das células somáticas híbridas:

- Fornecem uma maneira de segregar porções definidas do genoma para estudos adicionais;
- São essencialmente imortais em cultura;
- As células híbridas podem replicar e manter o segmento do genoma de interesse “retido”;
- As células híbridas podem ser isoladas com diferentes porções do genoma de interesse.

Estratégias Aplicáveis ao Mapeamento do Genoma de animais domésticos

#### 1 - Mapa físico

- 1.1. Genética de células somáticas
- 1.2. Hibridização “in situ”
- 1.3. Bibliotecas de DNA genômico (mapas “clone contig”)
- 1.4. Mapas de restrição (uso de enzimas de restrição)

#### 2 - Mapa genético

- 2.1. Animais híbridos interespecíficos
- 2.2. Células somáticas híbridas irradiadas

### ***Vantagens do método RH de mapeamento de genomas***

- Os marcadores não precisam ser polimórficos para serem mapeados.
- A contagem dos marcadores depende da ausência ou presença da banda e não na decifração de alelos.
- As distâncias físicas relativas ao longo do cromossomo são provavelmente menos distorcidas, pois nenhum "hot spot" de radio sensibilidade tem sido identificado com os conjuntos de células RH.
- A taxa de quebra do cromossomo nas linhagens RH é determinada pela quantidade de radiação usada, sendo portanto uma variável controlada.
- Um número ilimitado de linhagens de célula RH pode ser produzido a partir de uma única linhagem de célula doadora.

**Alguns sites para consulta:**

<http://bos.cvm.tamu.edu/>

<http://locus.jouy.inra.fr/cgi-bin/bovmap/intro.pl>

<http://www.ri.bbsrc.ac.uk>

## **Delineamentos experimentais para o mapeamento de QTL em animais**

Mário Luiz Martinez

Marcos Vinícius G. B. da Silva

Embrapa Gado de Leite

### **1. Introdução**

Grande número de delineamentos experimentais e de metodologias estatísticas tem sido proposto para se detectar genes que afetam características quantitativas com o auxílio de marcadores genéticos. Todos os delineamentos sugeridos têm vários elementos em comum. Considera-se que um *loco* de característica quantitativa (QTL) esteja ligado ao *loco* de um marcador com uma freqüência de recombinação  $r$ . A priori, pressupõe-se que apenas dois alelos estão segregando na população para ambos, o *loco* do marcador,  $M$ , e o *loco* do QTL,  $Q$ . Os genótipos do marcador serão denominados de  $MM$ ,  $Mm$  e  $mm$ . Os genótipos do QTL serão denominados por  $QQ$ ,  $Qq$  e  $qq$ , com efeitos esperados de  $a$ ,  $d$  e  $-a$ , respectivamente, sobre a característica quantitativa. Assim, se o indivíduo for heterozigoto para ambos os *locos*, metade da progênie receberá o alelo  $M$  e a outra metade o alelo  $m$ . Desde que  $M$  e  $Q$  estejam ligados, aquelas progênies que receberem  $M$  também receberão  $Q$ , enquanto aquelas que receberem  $m$  também receberão  $q$ , exceto pelos indivíduos recombinantes que receberão  $M$  com  $q$  ou  $m$  com  $Q$ . Dessa forma, o efeito do QTL pode ser detectado por meio da comparação das médias dos grupos de progênies que receberem alelos marcadores alternativos de pais heterozigotos. Os diversos delineamentos propostos diferem nos métodos utilizados para se "criar" pais heterozigotos para ambos os *locos* e nos cruzamentos realizados.

### **2. Delineamento experimental para detecção de QTL em cruzamentos de linhagens consangüíneas**

A maioria das análises realizadas para se detectar QTL tem sido baseada em cruzamentos planejados, embora alguns estudos tenham sido realizados com populações de animais existentes, especialmente gado de leite. O caso mais simples é o do cruzamento entre linhagens consangüíneas. Como mencionado anteriormente, o primeiro passo é cruzar duas linhagens em progênies heterozigotas  $F_1$ . Após esse passo, as progênies dos seguintes delineamentos têm sido consideradas para análises:

- i) Retrocruzamento (BC) de indivíduos  $F_1$  com uma das linhagens parentais;

- ii) Indivíduos F2 produzidos pelo cruzamento entre indivíduos F1 ou autofecundação de F1;
- iii) Linhagens consangüíneas recombinantes (RILF) produzidas pela autofecundação de indivíduos F2, ou pelo acasalamento de irmão com irmãs, nos casos em que a autofecundação não é possível (RILF);
- iv) Linhagens consangüíneas recombinantes (RILB) produzidas com indivíduos retrocruzados, ou acasalamentos de irmãos com irmãs (RILB);
- v) Linhagens duplas de haplóides (DH) produzidas pela autofecundação de duplos haplóides derivados de F1;
- vi) Progênies (P) produzidas pelo acasalamento de indivíduos F1 com uma terceira linhagem consangüínea.

Existem outros delineamentos mais complicados, mas todos são variações dos delineamentos acima listados.

A escolha do delineamento a ser usado deve levar em consideração os aspectos biológicos, econômicos, genéticos e estatísticos. Considerações biológicas estão relacionadas com o fato de que não são todos os delineamentos que podem ser usados com todas as espécies, como, por exemplo, o DH em animais domésticos. Para certas espécies, em que a autofecundação é a forma mais comum de reprodução, é mais fácil produzir grande número de indivíduos F2 do que BC ou P que requerem fertilização cruzada; para espécies em que os acasalamentos só ocorrem dentro da população, a consangüinidade pode resultar na redução do valor adaptativo em razão da presença de genes deletérios recessivos. Semelhantemente, o valor econômico dos delineamentos que são possíveis pode ser bem diferente. Por exemplo, se o objetivo é a introgressão de genes de uma característica específica presente em uma espécie selvagem ou nativa em uma outra previamente selecionada, o BC com a espécie selecionada terá valor econômico muito maior do que o F2. As considerações genéticas estão associadas com os parâmetros genéticos que desejamos estimar. Assim, os delineamentos BC ou P não permitem estimar relações de dominância, mas já o F2, sim. As considerações estatísticas estão relacionadas com os delineamentos que maximizem o poder de detectar os efeitos do QTL dentro de determinada situação. Por exemplo, nos delineamentos RI e DH todos os indivíduos dentro de uma linhagem terão o mesmo genótipo. Assim, será necessário genotipar apenas um único indivíduo de cada linhagem, enquanto o desempenho fenotípico de todos os indivíduos pode ser utilizado para determinar o efeito do QTL. Dessa forma, o poder estatístico será maior por indivíduo genotipado.

### 3. Delineamento de retrocruzamento (BC)

Na Figura 1 está ilustrado o delineamento de retrocruzamento com uma das raças parentais. A produção de F1 é obtida por meio do acasalamento entre duas linhagens que diferem em genótipos do QTL e do marcador. Normalmente, pressupõe-se que as duas linhagens parentais são homocigotas para os alelos alternativos de ambos os *locos*. Dessa forma, todos os indivíduos F1 terão o mesmo genótipo heterocigoto. Estes indivíduos F1 são então acasalados com uma das linhagens parentais. A composição genética destes indivíduos é de  $\frac{3}{4}$  do pai recorrente e  $\frac{1}{4}$  do outro pai. As progênes BC são então divididas em dois grupos baseados nos genótipos dos marcadores. À semelhança do que será considerado nos outros delineamentos experimentais, todos os demais *locos* que não estiverem ligados ao marcador genético em estudo estarão distribuídos aleatoriamente entre os grupos dos genótipos marcadores. Quando apenas um marcador é considerado, haverá apenas dois grupos genéticos marcadores para este tipo de delineamento e apenas um contraste poderá ser testado. Este contraste é a diferença entre as médias dos dois grupos de progênie que têm genótipos marcadores MM e mm.

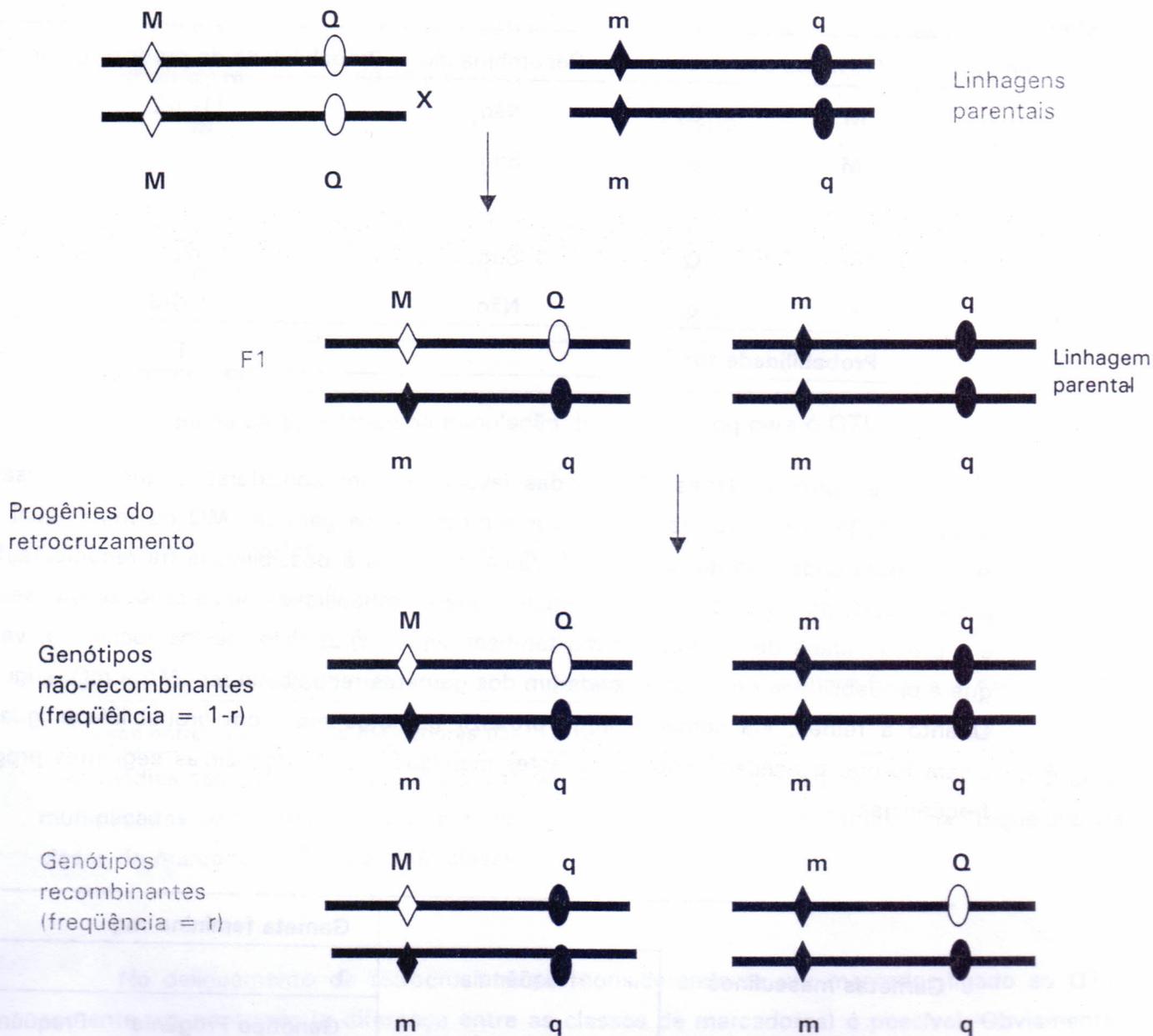


Figura 1. Delineamento de retrocruzamento (BC).

Dado que a frequência de recombinação é igual a  $r$ , devem ser calculadas as frequências gaméticas e genóticas para ambos os *locos*.

Em relação às frequências gaméticas, pode ser observado que o pai F1 é heterozigoto para ambos os *locos* (MmQq), o que o torna capaz de produzir quatro tipos de gametas:

Gametas possíveis		Recombinação	Probabilidade de produzir gametas
M	Q	Não	$(1-r)/2$
M	q	Sim	$r/2$
m	Q	Sim	$r/2$
M	q	Não	$(1-r)/2$
Probabilidade total			1

Estas probabilidades são obtidas levando-se em consideração que, na ausência de recombinação, a probabilidade de haver a produção de gametas **MQ** ou **mq** é igual a 1 (ou seja  $\frac{1}{2}$  para cada tipo de gameta). Todavia, como há a possibilidade de recombinação, esta probabilidade passa a ser **1-r**, o que implica que a probabilidade de se produzir **MQ** seja  $(1-r)/2$  e a probabilidade de se produzir **mq** também seja  $(1-r)/2$ . Pelo mesmo raciocínio, verifica-se que a probabilidade de se obter cada um dos gametas recombinantes (**Mq** e **mQ**) é igual a  $r/2$ . Quanto à fêmea, ela somente pode produzir gametas **mq**, com probabilidade igual a um. Desta forma, o acasalamento entre estes indivíduos pode produzir as seguintes progêneses e frequências:

Gametas masculinos	frequência	Gameta feminino mq	
		1	
		Genótipo Progênie	Frequência
<b>MQ</b>	$(1-r)/2$	<b>MmQq</b>	$(1-r)/2$
<b>Mq</b>	$r/2$	<b>Mmqq</b>	$r/2$
<b>mQ</b>	$r/2$	<b>mmQq</b>	$r/2$
<b>mq</b>	$(1-r)/2$	<b>mmqq</b>	$(1-r)/2$

Agora, considere-se, na progênie, que os indivíduos portadores do alelo **Q** para o QT possuam superioridade fenotípica igual a  $\alpha$  (efeito médio de substituição gênica, definido por Falconer, 1981) para a característica de interesse, em relação ao alelo **q**:

Alelo marcador obtido do touro	Alelo QTL obtido do touro	Freqüência	Média esperada na progênie <sup>1</sup>
M	Q	(1-r)/2	$\mu + \alpha$
M	q	r/2	$\mu$
m	Q	r/2	$\mu + \alpha$
m	q	(1-r)/2	$\mu$
<b>Freqüência total</b>		<b>1</b>	

<sup>1</sup>  $\mu$  = média da característica na população de genótipo qq para o QTL.

Espera-se, então, que as médias referentes a cada classe de marcador, sejam:

Média da classe do marcador M <sup>2</sup>	$[(1-r) \cdot (\mu + \alpha)/2 + (r \cdot \mu)/2] / 0,5 = \mu + (1-r) \cdot \alpha$
--	---

Média da classe do marcador m <sup>2</sup>	$[r \cdot (\mu + \alpha)/2 + (1-r) \cdot \mu/2] / 0,5 = \mu + r \cdot \alpha$
--	---

Diferença entre as classes de marcadores (D)	$(1-2r) \cdot \alpha$
--	-----------------------

<sup>2</sup> As médias são calculadas por meio do somatório das freqüência dos gametas ((1-r)/2 ou r) multiplicadas pelas médias do grupo (( $\mu + \alpha$ ) ou  $\mu$ ) e, então, divididas pela freqüência da classe do marcador (0,5 para cada classe).

No delineamento de retrocruzamento, considerando-se um marcador ligado ao QTL, somente um contraste (a diferença entre as classes de marcadores) é possível. Obviamente, se a diferença (dentro da família do touro) entre as classes de marcadores for significativa, testada pelo teste t ou pelo teste F, pode-se inferir que existe ligação entre marcador e QTL, sendo esta a base para detecção de QTL pela ANOVA ou por regressão, as quais a literatura trata como aproximação por **modelos lineares**. A vantagem principal deste modelo linear de análise é que ele pode ser executado pela maioria dos pacotes estatísticos existentes. As desvantagens são:

- 1) O efeito estimado está confundido com o efeito da recombinação entre o QTL e o marcador, e conseqüentemente a estimativa do efeito do QTL é viesada;
- 2) Pressupõe-se que os resíduos seguem distribuição normal;
- 3) Pressupõe-se que apenas marcador, bloco e erro aleatório influenciam a média da característica. Portanto, o método não é apropriado para múltiplos marcadores que estão ligados, que são correlacionados;

4) O método não diferencia entre o efeito de um QTL ligado e o efeito pleiotrópico do marcador genético.

#### 4. Delineamento de F2

Na Figura 2 está ilustrado o delineamento de F2. À semelhança do delineamento BC, dois pais homocigotos são cruzados para se obter o indivíduo heterocigoto F1. As progênies F2 são produzidas por meio da autofecundação dos indivíduos F1 ou pelo cruzamento entre eles. Como resultado, as progênies F2 serão de um dos três genótipos do marcador. Tais progênies são resultantes da união dos gametas não-recombinantes e recombinantes produzidos por machos e fêmeas F1, com as seguintes freqüências:

Gametas possíveis		Recombinação	Probabilidade de produzir gametas
M	Q	Não	$(1-r)/2$
M	q	Sim	$r/2$
M	Q	Sim	$r/2$
M	q	Não	$(1-r)/2$
Probabilidade total			1

As combinações de tais gametas podem produzir os genótipos e as freqüências apresentados na Tabela 2.

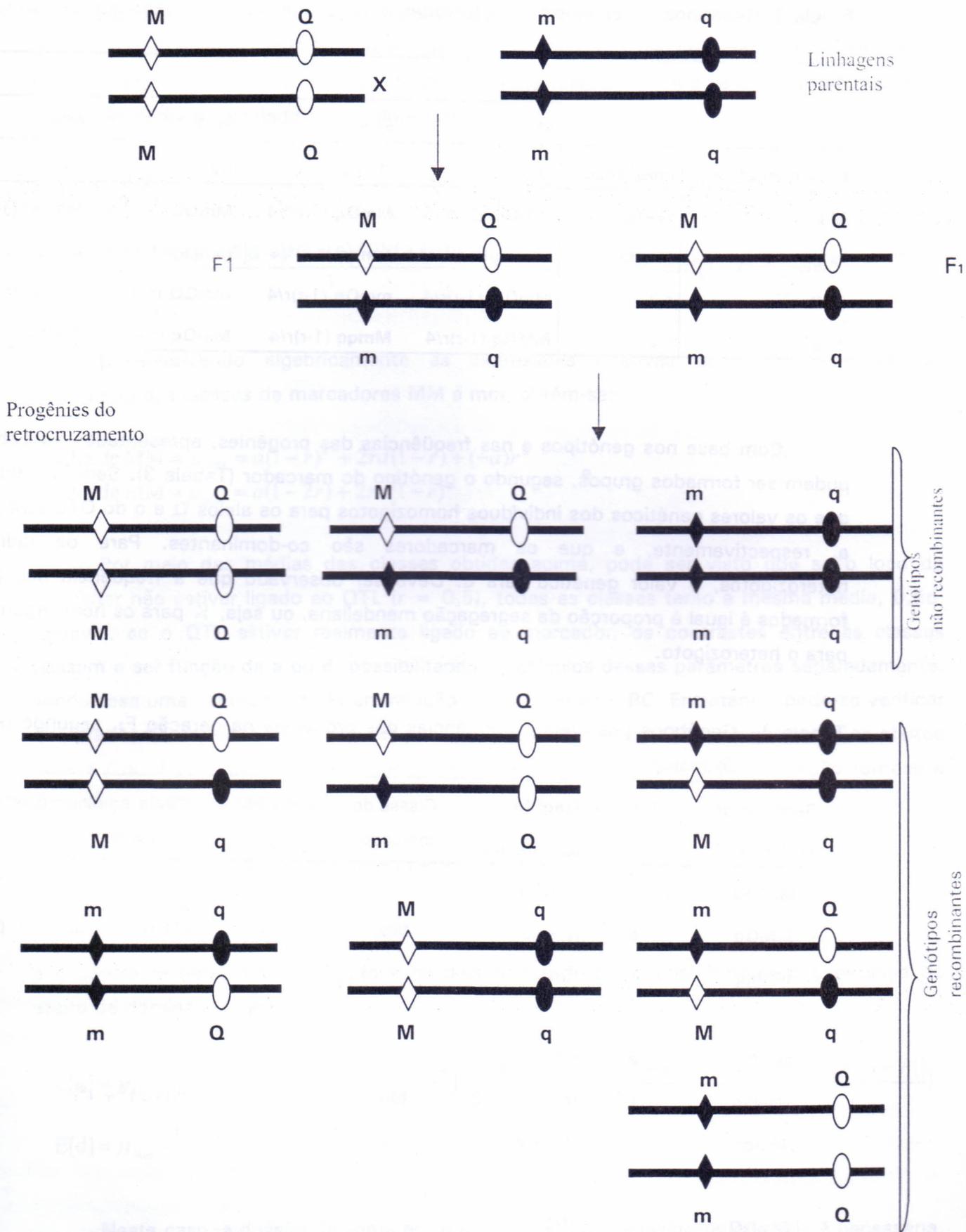


Figura 2. Delineamento F2.

Tabela 2. Genótipos e freqüências dos gametas e das progênies nas gerações F1 e F2.

		Óvulos			
		M Q	m q	m Q	M q
Espermatozóides	Freqüência	$(1-r)/2$	$(1-r)/2$	$r/2$	$r/2$
M Q	$(1-r)/2$	MMQQ $(1-r)^2/4$	MmQq $(1-r)^2/4$	MmQQ $(1-r)r/4$	MMQq $(1-r)r/4$
M q	$(1-r)/2$	MmQq $(1-r)^2/4$	mmqq $(1-r)^2/4$	mmQq $(1-r)r/4$	Mmqq $(1-r)r/4$
M Q	$r/2$	MmQQ $(1-r)r/4$	mmQq $(1-r)r/4$	mmQQ $r^2/4$	MmQq $r^2/4$
M q	$r/2$	MMQq $(1-r)r/4$	Mmqq $(1-r)r/4$	MmQq $r^2/4$	MMqq $r^2/4$

Com base nos genótipos e nas freqüências das progênies, apresentados na Tabela 2, podem ser formados grupos, segundo o genótipo do marcador (Tabela 3). Será admitido aqui que os valores genéticos dos indivíduos homocigotos para os alelos Q e q do QTL serão a e -a, respectivamente, e que os marcadores são co-dominantes. Para os indivíduos heterocigotos, o valor genético será d. Deve ser observado que a freqüência dos grupos formados é igual à proporção da segregação mendeliana, ou seja,  $\frac{1}{4}$  para os homocigotos e  $\frac{1}{2}$  para o heterocigoto.

Tabela 3. Genótipos possíveis e freqüências das progênies na geração F<sub>2</sub>, segundo o grupo do marcador.

Genótipo da progênie F <sub>2</sub>	Valor	Freqüência do genótipo	Classe do marcador	Freqüência do marcador	Média da classe do marcador
MMQQ	a	$(1-r)^2/4$	MM	$\frac{1}{4}$	$a(1-r)^2 + 2rd(1-r) + (-a)r^2$
MMQq	d	$r(1-r)/2$			
MMqq	-a	$r^2/4$			
MmQQ	a	$r(1-r)/2$	Mm	$\frac{1}{2}$	$d[(1-r)^2 + r^2]$
MmQq	d	$[r^2 + (1-r)^2]/2$			
Mmqq	-a	$r(1-r)/2$			
MmQQ	a	$r^2/4$	Mm	$\frac{1}{4}$	$(-a)(1-r)^2 + 2rd(1-r) + ar^2$
MmQq	d	$r(1-r)/2$			
Mmqq	-a	$(1-r)^2/4$			

As médias, dentro de cada classe do marcador, são obtidas pelo somatório dos produtos dos valores genéticos relativos a cada genótipo de QTL por suas respectivas frequências, dividido pela frequência da classe. Como exemplo, tome-se a classe do marcador **MM**, cuja média é calculada da seguinte forma:

Média de **MM**

$$= \mu_{MM} = \frac{[(1-r)^2/4]a + [r(1-r)/2]d + (r^2/4)(-a)}{1/4} = a(1-r)^2 + 2rd(1-r) + (-a)r^2.$$

Desenvolvendo algebricamente as expressões relativas às médias (ou valores esperados) das classes de marcadores **MM** e **mm**, obtém-se:

$$\begin{aligned} \text{Média de MM} = \mu_{MM} &= a(1-r)^2 + 2rd(1-r) + (-a)r^2 \\ \text{Média de mm} = \mu_{mm} &= a(1-2r) + 2rd(1-r) \end{aligned} \quad e$$

Por meio das médias das classes obtidas acima, pode ser visto que se o loco do marcador não estiver ligado ao QTL ( $r = 0,5$ ), todas as classes terão a mesma média, **0,5d**. Todavia, se o QTL estiver realmente ligado ao marcador, os contrastes entre as classes passam a ser função de **a** ou **d**, possibilitando os cálculos desses parâmetros separadamente, sendo essa uma vantagem do  $F_2$  em relação ao delineamento BC. Entretanto, pode-se verificar que praticamente todos os contrastes possíveis incluem o termo  $(1-2r)$ , o que torna os efeitos de **a** e **r** ou **d** e **r** confundidos. O contraste entre os homozigotos, cuja estimativa fornece a Esperança do efeito aditivo, é:

$$\begin{aligned} E[a] &= (\mu_{MM} - \mu_{mm}) = a(1-2r) + 2rd(1-r) - [-a(1-2r) + 2rd(1-r)] \\ E[a] &= (\mu_{MM} - \mu_{mm}) = 2a(1-2r) \end{aligned}$$

e o contraste entre o heterozigoto e os dois homozigotos, o qual fornece a Esperança do efeito de dominância, é:

$$\begin{aligned} E[d] &= \mu_{Mm} - \frac{(\mu_{MM} + \mu_{mm})}{2} = d[(1-r)^2 + r^2] - \left\{ \frac{a(1-2r) + 2rd(1-r) + [-a(1-2r) + 2rd(1-r)]}{2} \right\} \\ E[d] &= \mu_{Mm} - \frac{(\mu_{MM} + \mu_{mm})}{2} = d(1-2r)^2 \end{aligned}$$

Neste caso, a divisão da soma entre as médias dos homozigotos por dois é necessária, de modo a atender as exigências matemáticas para o estabelecimento do contraste. Aqui são apresentados somente dois contrastes, todavia, outros podem ser estabelecidos, de acordo com o interesse do pesquisador, sendo a significância testada pelo teste de **t** ou **F**.

Entretanto, segundo Martinez (2001), embora se possa testar para os efeitos aditivos (a) e de dominância (d), ambas as estimativas não podem ser obtidas sem viés, por meio de modelos lineares.

Utilizando as freqüências gaméticas contidas na Tabela 2, podem ser obtidas as probabilidades conjuntas e marginais e, conseqüentemente, as condicionais. Assim, a probabilidade conjunta de **MQ** é a probabilidade de que um indivíduo  $F_2$  tenha o genótipo **MMQQ**, que é igual a  $P(\mathbf{MQ})$ , sendo  $P(\mathbf{MQ}) = [(1-r)/2] \cdot [(1-r)/2] = (1-r)^2/4$ . A probabilidade marginal de **MM**, ou seja, a probabilidade de que um indivíduo  $F_2$  tenha o genótipo **MM** é  $1/4$ . Usando a fórmula de probabilidade condicional, tem-se que:

$$P(Q_k | M_j) = \frac{P(Q_k M_j)}{P(M_j)} = \frac{(1-r)^2 / 4}{1/4} = (1-r)^2.$$

As demais probabilidades apresentadas a seguir são obtidas da mesma maneira.

$P(\mathbf{QQ}   \mathbf{MM}) = (1-r)^2$	$P(\mathbf{Qq}   \mathbf{MM}) = 2r(1-r)$	$P(\mathbf{qq}   \mathbf{MM}) = r^2$
$P(\mathbf{QQ}   \mathbf{Mm}) = r(1-r)$	$P(\mathbf{Qq}   \mathbf{Mm}) = (1-r)^2 + r^2$	$P(\mathbf{qq}   \mathbf{Mm}) = r(1-r)$
$P(\mathbf{QQ}   \mathbf{mm}) = r^2$	$P(\mathbf{Qq}   \mathbf{mm}) = 2r(1-r)$	$P(\mathbf{qq}   \mathbf{mm}) = (1-r)^2$

Por meio das probabilidades condicionais, podem ser calculadas as médias para cada classe de marcador, bem como os contrastes de interesse, do mesmo modo que no item anterior.

## 5. Delineamentos experimentais grandes populações segregantes

Para muitas espécies de animais domésticos, árvores frutíferas perenes e, especialmente, em seres humanos, a formação de linhagens endogâmicas, as quais formam a base dos delineamentos descritos anteriormente, é impraticável. Dessa forma, uma alternativa é a execução de análises em famílias existentes dentro da população. Todavia, como os experimentos para o mapeamento de QTLs envolvem a genotipagem e a mensuração fenotípica de grande número de indivíduos, 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), as populações exogâmicas (*outbred populations*) podem ser de interesse por outras razões, que não as restrições experimentais. Segundo esses 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. 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 resulta em significativa diminuição do poder de detecção. Com o uso de delineamentos envolvendo linhagens endogâmicas, todos os pais F1 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 avós) será heterozigota para o QTL e para o marcador, sendo a fase de ligação desconhecida. Nesse caso, os indivíduos poderão diferir quanto à fase de ligação. Assim, o alelo **M** para o marcador poderá estar ligado ao alelo **Q** do QTL em alguns pais, e com o alelo **q** em outros, o que demonstra a necessidade de se efetuar as análises considerando-se os pais individualmente. Além disso, na população F1 haverá somente dois alelos segregando, enquanto um número indefinido de alelos pode estar segregando na população exogâmica.

Outro ponto importante, o qual está ligado ao poder de detecção, é que os efeitos dos QTLs são 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.

### 5.1 Delineamento de filhas

Weller et al. (1990) relataram que o mapeamento de QTLs por meio de marcadores é baseado na esperança de que, se um indivíduo é heterozigoto para o loco do marcador e do QTL, então, a menos que ocorra recombinação, a progênie que recebe 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 ligado tenderá a gerar diferença entre as médias dos dois grupos, em relação à característica quantitativa de interesse, a qual, se for significativa, indicará a presença do QTL próximo ao marcador.

Baseados nesse princípio, Neimann-Sorensen e Robertson (1961) propuseram o **delineamento de filhas** (*daughter design*), o qual é usado principalmente em gado de leite, em que grande número de progênies por touro pode ser obtido, 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. Na Figura 3 é ilustrado o delineamento de filhas para o caso de uma única família, sendo que a letra **x**, associada aos alelos do marcador e do QTL, indica o desconhecimento dos alelos presentes na população de vacas.

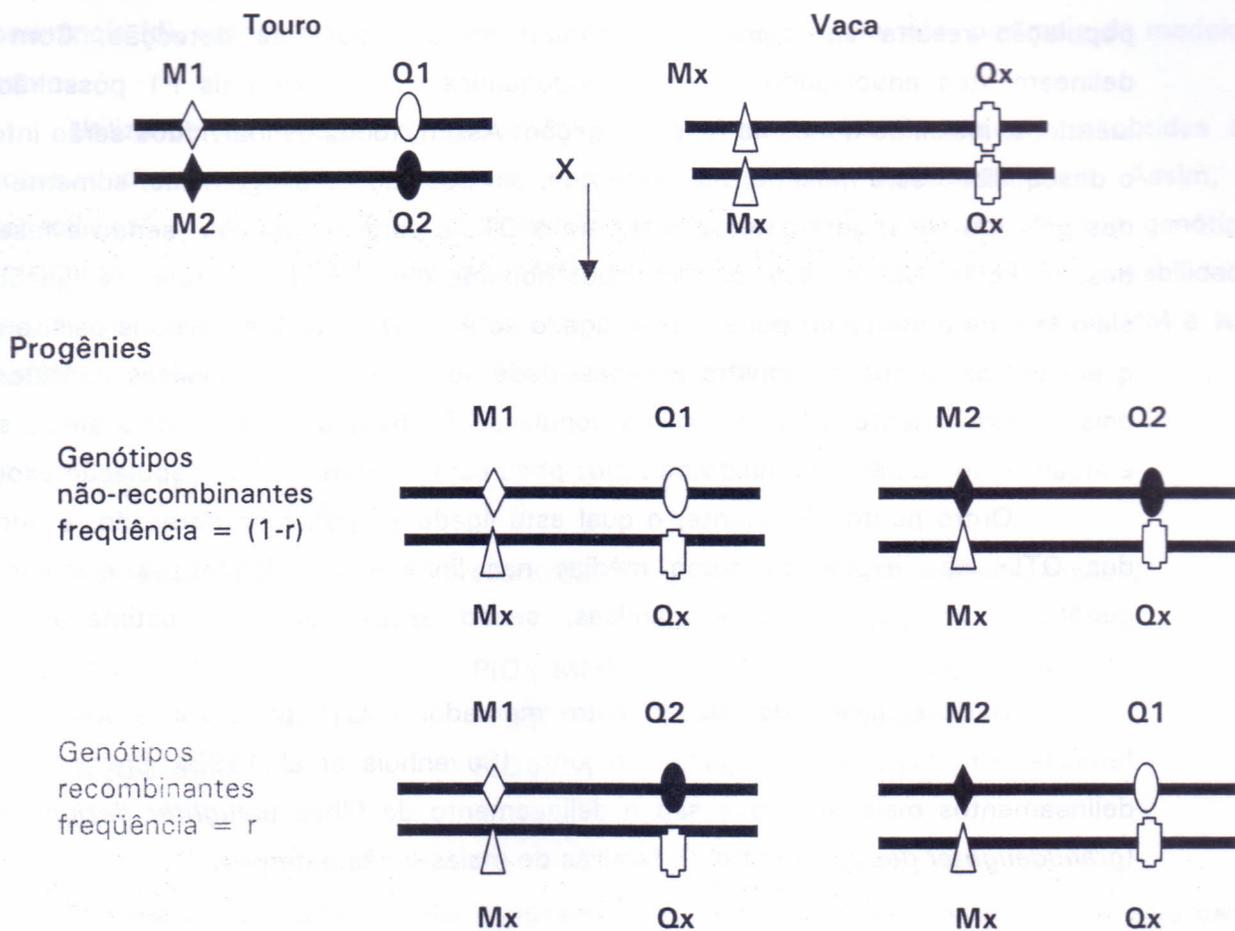


Figura 3. Delineamento de Filhas.

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, o que pode tornar o experimento mais caro. Duas situações são mais comuns, a primeira em que são analisados os registros de meias-irmãs, filhas de um único touro, e a segunda, em que o conjunto de dados disponível refere-se às famílias de meias-irmãs, filhas de vários touros. No caso em que temos apenas um touro com várias filhas, a análise pode ser pelo seguinte modelo:

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

em que

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

$\mu$  = 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. Valor significativo para o teste revela a ligação entre marcador e QTL, o qual é heterozigoto no touro.

Nos casos em que há limitado número de filhas por reprodutor, a análise de conjuntos de dados contendo informações de várias famílias de touros aumentará o poder de detecção. Nesta situação, o modelo apropriado será:

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

em que

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

$\mu$  = 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  $M_1$  será comparado com o grupo de progênie que recebeu o alelo  $M_2$ , dentro da família de cada touro.

## 5.2 Delineamento de netas

Hipótese alternativa ao delineamento de filhas é o chamado **delineamento de netas** (Figura 4), proposto por Weller et al. (1990) com o objetivo de aumentar o poder de detecção. Neste delineamento, touros heterozigotos (aqui chamados de avós) têm os filhos genotipados para loco do marcador e suas netas (filhas de seus filhos) avaliadas para as características fenotípicas. Pressupõe-se que as vacas são acasaladas ao acaso com os touros heterozigotos e com seus filhos. Dessa forma, segundo Weller et al. (1990), apenas a metade das netas receberá o alelo paterno, e a esperança do contraste entre os grupos de progênies (netas) é apenas a metade do que é no delineamento de filhas. Entretanto, segundo esses autores, o maior número de registros de desempenho avaliado pode compensar a redução do valor do contraste, diminuindo seus respectivos erros-padrão. Outra vantagem é a diminuição do custo do programa de avaliação, pois menor número de animais é genotipado, além de facilitar a colheita de amostras de sangue ou sêmen dos filhos do touro, os quais geralmente estarão concentrados em centrais de inseminação artificial, enquanto suas filhas estarão espalhadas por diversas fazendas.

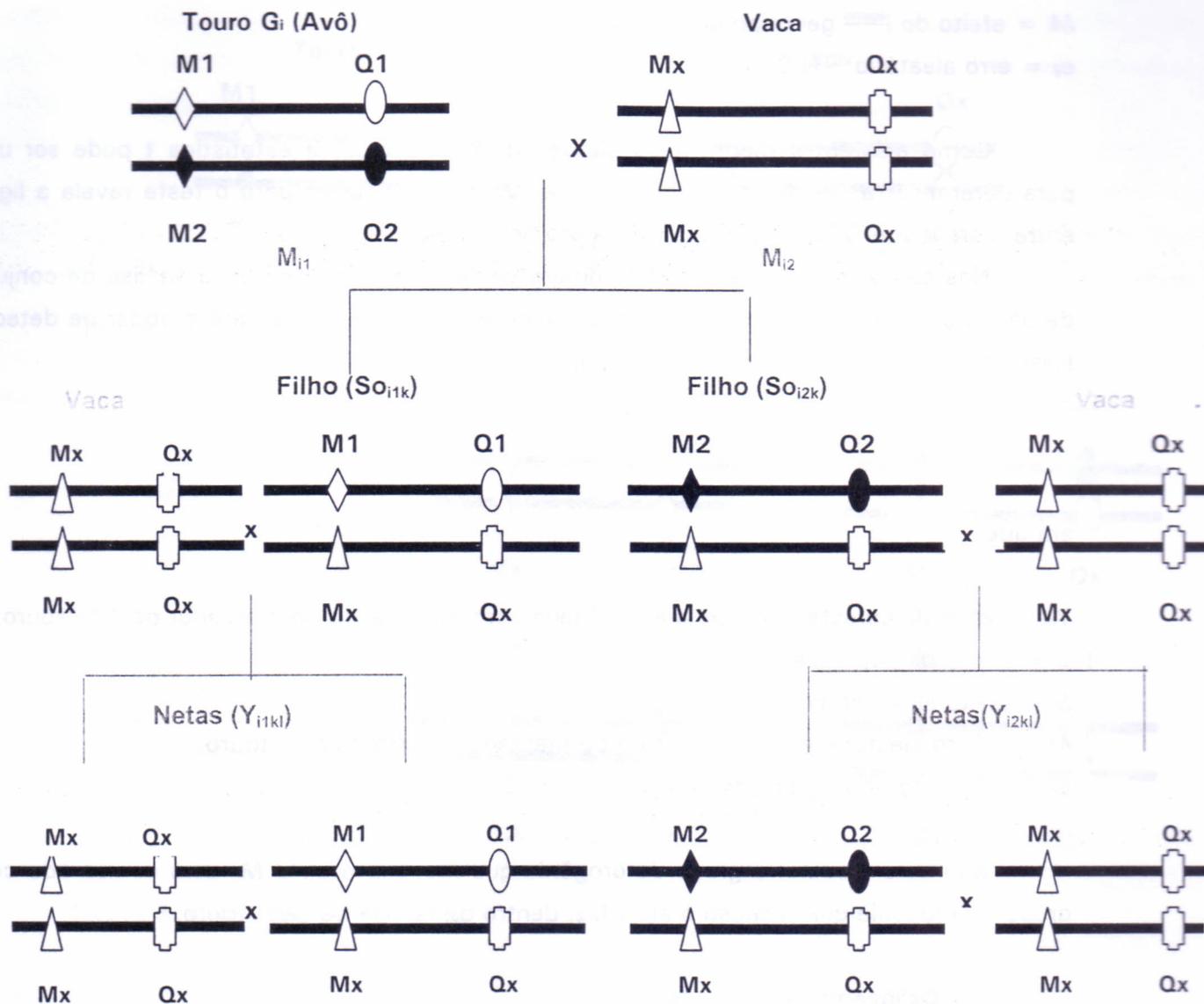


Figura 4. Delineamento de netas.

Neste caso, o modelo pode ser:

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

em que

$Y_{ijkl}$  = valor da característica para a  $i^{\text{ésimo}}$  neta, filha do  $k^{\text{ésimo}}$  reprodutor (filho) com o  $j^{\text{ésimo}}$  alelo do marcador do  $i^{\text{ésimo}}$  touro (avô);

$\mu$  = média da população;

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

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

$SO_{ijk}$  = efeito do  $k^{\text{ésimo}}$  reprodutor (filho) com o  $j^{\text{ésimo}}$  alelo do marcador, filho do  $i^{\text{ésimo}}$  touro (avô);

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

## 6. Delineamento para detecção de QTL em famílias pequenas de populações segregantes

Haseman e Elston propuseram o método de análise de par de irmãos em 1972. Este método é mais apropriado à população de humanos, na qual famílias pequenas de irmãos-completos são mais comuns. O delineamento básico é mostrado na Figura 5. Na situação mais simples, os dois pais têm quatro alelos marcadores diferentes: M1, M2, M3 e M4. Pressupõe-se também que este *loco* está ligado a um QTL heterozigoto para ambos os pais. Na figura, são apresentados apenas dois tipos de par de irmãos. Para efeito de simplicidade, apenas os não-recombinantes são apresentados.

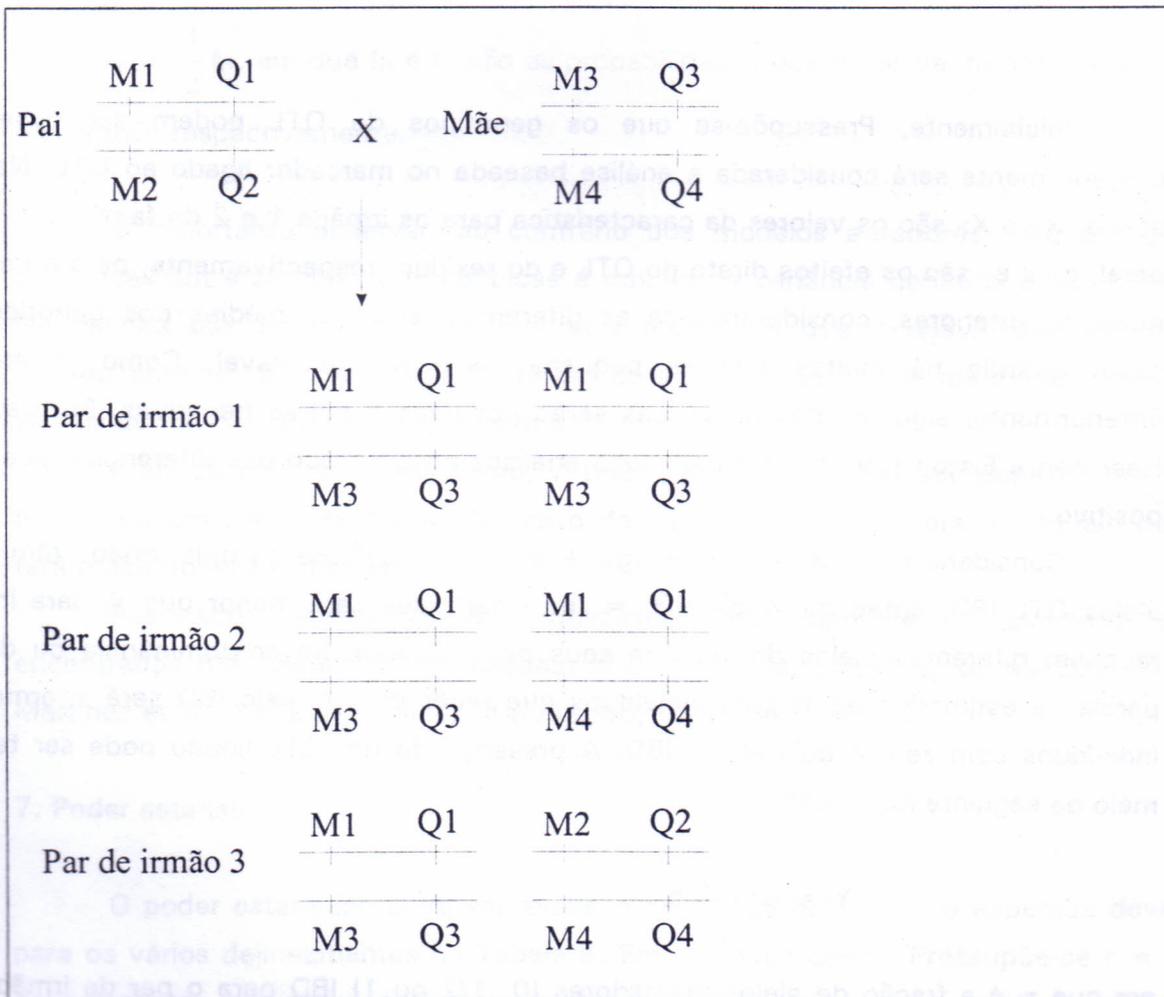


Figura 5. Delineamento de par de irmãos-completos.

No par de irmãos 1, os dois alelos marcadores de ambos os irmãos são idênticos por descendência (IBD). Ambos os irmãos receberam os mesmos dois alelos marcadores de ambos os pais. No par de irmãos 2, ambos os irmãos receberam o alelo M1 do pai, mas alelos diferentes da mãe. Assim, um par de alelos é IBD, enquanto o outro não é. No par de irmãos 3, nenhum dos alelos marcadores é IBD. Pressupondo-se que o *loco* marcador está ligado ao QTL, como mostrado na Figura 5, os indivíduos com mais *locos* marcadores IBD deverá também ser mais similar para a característica quantitativa. À semelhança do caso de grandes famílias de meio-irmãos, a análise deve levar em consideração a estrutura da família, ou seja, um QTL ligado ao marcador pode não resultar em efeito geral do genótipo marcador. Haseman e Elston (1972) propuseram o seguinte modelo estatístico:

$$X_{1j} = \mu + g_{1j} + e_{1j} \quad (3)$$

$$X_{2j} = \mu + g_{2j} + e_{2j}$$

Inicialmente, Pressupõe-se que os genótipos do QTL podem ser observados e posteriormente será considerada a análise baseada no marcador ligado ao QTL. Na equação acima,  $X_{1j}$  e  $X_{2j}$  são os valores da característica para os irmãos 1 e 2 da família  $j$ ,  $\mu$  é a média geral,  $g_{1j}$  e  $e_{1j}$  são os efeitos direto do QTL e do resíduo, respectivamente, para o irmão  $i$ . Nos modelos anteriores, consideraram-se as diferenças entre as médias dos genótipos. Neste caso, quando há muitas famílias pequenas isso não é viável. Como já mencionado anteriormente, algumas das diferenças serão positivas e outras negativas. Por causa disso, Haseman e Elston (1972) basearam suas análises no quadrado das diferenças, que é sempre positivo.

Considere  $Y_i = (X_{1j} - X_{2j})^2 = (g_{1j} + e_{1j} - g_{2j} - e_{2j})^2$ . Se os dois irmãos têm ambos os alelos QTL IBD, então  $g_{1j} = g_{2j}$  e  $Y_j = (e_{1j} - e_{2j})^2$ , que será menor que  $Y_j$  para irmãos que recebam diferentes alelos do QTL de seus pais. Quando há co-dominância ou dominância parcial, a estimativa de  $Y_j$  para indivíduos que recebam um alelo IBD será intermediária em indivíduos com zero e dois alelos IBD. A presença de um QTL ligado pode ser testada por meio da seguinte regressão:

$$Y_j = \alpha + \beta \pi_j, \quad (4)$$

em que  $\pi_j$  é a fração de alelos marcadores (0, 1/2 ou 1) IBD para o par de irmãos  $j$ ;  $\alpha$  é o ponto de interseção no eixo  $y$ ; e  $\beta$  é o coeficiente de regressão, que será sempre negativo. A detecção do QTL ligado ao marcador pode ser testada por meio do valor de  $\beta$ . Mesmo no caso de ligação incompleta entre QTL e marcador,  $\beta$  ainda será negativo. No caso de não

existir ligação (hipótese de nulidade),  $\beta$  será igual a zero. A estimativa de  $\beta$  dependerá de vários fatores, incluindo a taxa de recombinação entre QTL e marcador, o número de alelos do QTL que estão segregando na população e a relação de dominância entre os alelos.

Para o caso de ligação parcial, sem dominância e informação completa dos pais, Haseman e Elston (1972) demonstraram que a esperança de  $\beta = -2(1 - 2r)^2 \sigma_a^2$ . É importante esclarecer que o termo "informação completa dos pais" significa dizer que o número de alelos marcadores IBD pode ser determinado para cada par de irmãos. Este é o caso ilustrado na Figura 4, em que os dois pais têm quatro alelos diferentes. Isto é verdade também para o caso em que ambos os pais são heterozigotos, mas tem um mesmo alelo em comum. Todavia, para outras combinações de alelos marcadores não será possível se determinar inequivocadamente o número de alelos IBD para todas as combinações possíveis de pares de irmãos. Para esses casos, Haseman e Elston (1972) demonstraram que  $\pi_j$  pode ser estimado por:

$\hat{\Pi}_j = f_{j2} + \frac{1}{2} f_{j1}$ , em que  $f_{j1}$  e  $f_{j2}$  são as probabilidades que o par de irmãos  $j$  tenha um ou dois alelos IBD, respectivamente.

É importante observar, ao contrário dos modelos anteriores, que a esperança das diferenças entre as classes genotípicas é função da variância genética e não mais de  $a$  e  $d$ . Na maioria das situações práticas  $a$  será menor do que o desvio padrão do resíduo e conseqüentemente o delineamento de par de irmãos pode ser de menor poder estatístico do que os outros delineamentos anteriormente mencionados.

Além do modelo de regressão,  $Y_j$  pode também ser analisado com o número de alelos IBD como um efeito de classe. No caso de dominância incompleta, o modelo de regressão terá maior poder estatístico.

Resultados com simulações, utilizando desse tipo de delineamento, podem ser encontrados nas seguintes publicações: Martinez et al. (1998 a, b), Martinez (1998 a, b), Martinez et al. (1999) e Martinez et al. (2000), Martinez (2001).

## 7. Poder estatístico

O poder estatístico pode ser avaliado com base na variância esperada devida ao QTL para os vários delineamentos na Tabela 5. Em todos os casos, Pressupõe-se  $r = 0$ . Para os delineamentos em populações segregantes, admitiram-se também apenas dois alelos com igual freqüência para o QTL. Os resultados para o F-2 foram calculados considerando-se  $d = a$  e  $d = 0$ . As variâncias para os demais delineamentos foram obtidas considerando-se  $d = 0$ .

Tabela 5. Variância esperada devida ao QTL.

Variância do QTL	Delineamentos					
	F-2 (d = a)	F-2 (d = 0)	BC	Irmãos completos	Meios-irmãos	Netas
	$3a^2/2$	$a^2/2$	$a^2/4$	$a^4/8$	$a^2/8$	$a^2/32$

Exceto pelo delineamento de irmãos completos, a variância devida ao QTL é função de  $a^2$ . Considerando que o efeito do QTL geralmente é menor do que o desvio padrão do resíduo, o delineamento de irmãos completos terá variância devida ao QTL muito menor do que os outros delineamentos. À exceção desse delineamento, os demais estão listados na Tabela 5, em ordem decrescente da variância do QTL. A variância do QTL é aproximadamente proporcional ao inverso do tamanho da amostra necessária para se ter determinado poder de detecção de QTL. Dessa forma, o delineamento de netas exigirá aproximadamente 16 vezes mais registros para se obter o mesmo poder que o delineamento F2 com  $d = 0$ .

Quando gerações adicionais são genotipadas e medidas para a característica quantitativa, o poder de detecção não aumenta por indivíduo genotipado, mas a precisão no mapeamento do QTL aumenta, em razão de a geração adicional de recombinação. A vantagem da geração adicional com relação ao mapeamento do QTL será considerado em maiores detalhes posteriormente.

## 8. Referências Bibliográficas

- BOVENHUIS, H.; VAN ARENDONK, J.A.M.; DAVIS, G.; et al. Detection and mapping of quantitative trait loci in farm animals, *Livest. Prod. Sci.*, 52: 135-144, 1997.
- HASEMAN, J.K; ELSTON, R.C. The investigation of linkage between a quantitative trait and a marker loco. *Behav. Genet.* 2: 3-19, 1972.
- LYNCH, M.; WALSH, B. *Genetics and analysis of quantitative traits*. 1. ed. Massachussets: Sinauer Associates, Inc., 1998. 980p.
- MARTINEZ, M.L.; Vukasinovic, N.; Freeman, A. E.; Fernando, R. L. Mapping QTL in outbred populations using selected samples. *Genetic Selection Evolution*, V. 30, p. 453-468. 1998a.

- MARTINEZ, M. L. Eficiência dos delineamentos de filhas e netas para detectar ligação entre marcador loci de característica quantitativa. *Revista Bras. de Zootec.* v. 27, n. 4, p. 707-714. 1998a.
- MARTINEZ, M. L. Efeitos da substituição de alelos, da frequência gênica e da taxa de recombinação sobre o poder de se detectar ligação entre marcador genético e loci de característica quantitativa de baixa heritabilidade. *Revista Bras. de Zootec.*, v. 27, p. 715-721, 1998b.
- MARTINEZ, M. L.; Vukasinovic, N.; Freeman, A. E. Estimating QTL location and QTL variance in half-sib families under the random model with missing parental genotypes. *Journal of Animal Breeding and Genetics.* V. 115, n. 3, p. 165-180. 1998b.
- MARTINEZ, M. L.; Vukasinovic, N.; Freeman, A. E. Random model approach formapping QTL in half-sib families. *Genetics Selection Evaluation*, v. 31, p. 319-340. 1999.
- MARTINEZ, M. L.; Vukasinovic, N. Algoritmo para cálculo da proporção de genes idênticos por descendência, para mapear QTL em famílias de meio-irmãos. *Revista Bras. de Zootec.*, v. 29, n.2, p. 443-451. 2000.
- MARTINEZ, M. L. Delineamentos experimentais para detecção de locos de características quantitativas (QTL) em animais. In *Biologia molecular aplicada à produção animal*. Brasília. Embrapa Informação Tecnológica, 2001. p. 62-78.
- MARTINEZ, M. L. Métodos estatísticos para a detecção de locos de características quantitativas. In: REGITANO, L.C.A. & COUTINHO L.H.(Ed.) *Biología molecular aplicada à produção animal*. 1.ed. Brasília: Embrapa Informação Tecnológica, 2001, p. 79-102.
- NEIMANN-SOESSEN, A.; ROBERTSON, A. The association between blood groups and several production characters in three Danish cattle breeds. *Acta. Agr. Scand.* 11: 163-196, 1961.
- WELLER, J.I., KASHI, Y. AND SOLLER, M. Power of "daughter" and "granddaughter" designs for genetic mapping of quantitative traits in dairy cattle using genetic markers. *J. Dairy Sci.* 73: 2525-2537, 1990.

## Métodos e modelos estatísticos usados no mapeamento de QTLs

Marcos Vinicius G. Barbosa da Silva

Mário Luiz Martinez

Embrapa Gado de Leite

### 1. Introdução

Basicamente, os delineamentos experimentais para identificação e mapeamento de QTL podem ser divididos em dois grupos: aqueles baseados no cruzamento entre linhagens ou raças, que diferem na característica de interesse, e os métodos baseados em populações segregantes.

A análise de registros provenientes de cruzamentos entre linhagens endogâmicas apresenta algumas vantagens, como o conhecimento dos genótipos parentais e da fase de ligação e do número de alelos do suposto QTL. Também, esse tipo de dado pode ser considerado como sendo de uma única grande família, porque todos os indivíduos compartilham o mesmo genótipo parental, o que resulta em estimação direta dos efeitos de substituição e de dominância do QTL (Zeng, 1994). Todavia, muitas vezes a implementação desses cruzamentos é impraticável e onerosa (Boilard e Boichard, 2002).

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

Atualmente, são conhecidos três métodos de análise: os métodos baseados em modelos fixos, em modelos mistos ou em modelos aleatórios. Esses métodos são categorizados, com base no modelo genético-aditivo adotado (número de alelos do QTL, 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 técnica de Monte Carlo em Cadeia de Markov (Bink e Van Arendonk, 2000), podem ser utilizados para estimação de médias e variâncias dos genótipos de QTL, bem como da taxa de recombinação ( $r$ ) entre os marcadores genéticos e o QTL.

Assim, os objetivos deste trabalho são abordar as aplicações, as vantagens e as desvantagens de cada um desses modelos.

## 2. Aplicação de modelos lineares fixos

No delineamento  $F_2$ , considerando-se um marcador ligado ao QTL, os contrastes entre as médias das classes são funções dos efeitos aditivo, ou  $a$  (contraste entre as médias dos homozigotos), e de dominância, ou  $d$  (contraste entre as médias do heterozigoto e dos dois homozigotos), possibilitando os cálculos desses parâmetros separadamente, sendo essa uma vantagem do  $F_2$  em relação ao delineamento de retrocruzamento (BC). Entretanto, os contrastes possíveis incluem o termo  $(1-2r)$ , o que torna os efeitos de  $a$  e  $r$  ou  $d$  e  $r$ , confundidos. Aqui são apresentados somente dois contrastes, todavia, outros podem ser estabelecidos, de acordo com o interesse do pesquisador, sendo a significância verificada pelo teste de  $t$  ou  $F$ . Entretanto, segundo Martinez (2001), embora se possa testar os efeitos aditivos e de dominância, ambas as estimativas não podem ser obtidas sem viés, por meio de modelos lineares. Adicionalmente, muitas vezes não se consegue distinguir grande efeito de QTL inapropriadamente marcado (alto  $\alpha^1$  e grande distância do marcador) ou pequeno efeito de QTL estreitamente ligado ao marcador (baixo  $\alpha$  e pequena distância do marcador). A seguir, serão descritos os métodos usados para mapeamento de QTL, por meio de modelos lineares, em populações constituídas por indivíduos  $F_2$ .

### 2.1 Análise de variância e teste $t$ usando a progênie $F_2$

Para a ANOVA, o modelo linear é

$$Y_{i(j)k} = \mu + M_i + g(M)_{j(i)} + e_{i(j)k},$$

em que

$Y_{i(j)k}$  = valor da característica para o  $j^{\text{ésimo}}$  indivíduo com o  $i^{\text{ésimo}}$  marcador na  $k^{\text{ésima}}$  repetição;

$\mu$  = média da população;

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

$g(M)_{j(i)}$  = efeito genotípico dentro do genótipo do marcador;

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

O termo  $g(M)_{j(i)}$  refere-se ao efeito genotípico de outros QTLs, que não estão ligados ao marcador. Ao se proceder a análise utilizando pacotes estatísticos, como o SAS (STATISTICAL ANALYSIS SYSTEM, 1990), esse efeito estará incluído no resíduo.

Dessa forma, o teste adequado é dado por

---

<sup>1</sup> Média de efeito de substituição gênica, definido por FALCONER e MACKAY (1996).

$$F = \frac{QMM}{QME}, \text{ associada a } 2 \text{ e } N-2 \text{ graus de liberdade.}$$

O quadrado médio de marcador pode ser decomposto em dois contrastes ortogonais, o primeiro referente ao efeito aditivo e o segundo ao efeito de dominância, sendo a significância de ambos testada pela estatística  $F$ , com  $QME$  no denominador, e associada a  $1$  e  $N-2$  GL. A estatística  $t$ , nesse caso, é dada por

$$t_M = \frac{\text{contraste de interesse}}{\sqrt{\hat{S}_M^2 \left( \frac{1}{n_1} + \frac{1}{n_2} \right)}}, \text{ com } N-2 \text{ graus de liberdade,}$$

em que  $n_1$  e  $n_2$  são os números de progênies para os genótipos de marcadores sendo testados e  $\hat{S}_M^2$  é a variância conjunta dentro das três classes de marcadores. Ressalte-se que quando houver apenas  $1$  grau de liberdade para o numerador da estatística  $F$ , esta é igual ao quadrado da estatística  $t$  ( $F = t^2$ ).

As hipóteses de nulidade testadas, com base nos contrastes, correspondem à ausência de efeitos genéticos relativos ao QTL ( $H_0: a = 0$  e  $d = 0$ ) ou a inexistência de ligação entre QTL e marcador ( $H_0: r = 0,5$ ).

Deve-se observar que, se o modelo usado fosse referente a marcadores dominantes, a esperança do contraste seria alterada e, nesse caso, não seria possível dissociar os efeitos aditivo e de dominância.

## 2.2 Regressão linear usando a progênie $F_2$

A regressão linear pode ser usada para a análise de um único marcador na progênie  $F_2$ . Entretanto, a existência dos efeitos aditivo e de dominância torna necessária a inclusão de mais uma variável independente no modelo, o qual se torna

$$Y_j = \beta_0 + \beta_1 X_{1j} + \beta_2 X_{2j} + \varepsilon_j,$$

em que

$Y_j$  = valor fenotípico da característica para o  $j^{\text{ésimo}}$  indivíduo na população;

$\beta_0$  = coeficiente linear da reta;

$\beta_1$  = coeficiente de regressão referente ao efeito aditivo;

$\beta_2$  = coeficiente de regressão referente ao efeito de dominância;

$$x_{1j} = \text{variável indicadora do efeito aditivo} \begin{cases} 1, \text{ se o indivíduo tem genótipo } MM \\ 0, \text{ se o indivíduo tem genótipo } Mm \\ -1, \text{ se o indivíduo tem genótipo } mm \end{cases}$$

$$x_{2j} = \text{variável indicadora do efeito de dominância} \begin{cases} -1, \text{ se o indivíduo tem genótipo } MM \\ 1, \text{ se o indivíduo tem genótipo } Mm \\ -1, \text{ se o indivíduo tem genótipo } mm \end{cases}$$

$\varepsilon_j$  = erro aleatório  $\sim$  NID  $(0, \sigma^2)$ .

Considerando-se que a frequência dos grupos formados é igual à proporção da segregação mendeliana, ou seja, 0,25 para os homocigotos e 0,5 para o heterocigoto, as esperanças dos parâmetros para esse modelo e para os valores das variáveis indicadoras  $x_{1j}$  e  $x_{2j}$  são apresentadas por Liu (1998).

Os testes de hipóteses são realizados com base no valor da soma de quadrados residual, considerando-se o modelo completo e um modelo reduzido do qual é retirado o parâmetro que se quer testar. Assim, os modelos reduzidos são obtidos de acordo com a hipótese que se quer testar.

Para testar  $H_0: a = 0$  (ausência do efeito aditivo), o modelo reduzido é

$$y_j = \beta_0 + \beta_2 x_{2j} + \varepsilon_j \text{ para } \beta_1 = 0,$$

enquanto para  $H_0: d = 0$  (ausência do efeito de dominância), o modelo adequado é

$$y_i = \beta_0 + \beta_1 x_{1j} + \varepsilon_j \text{ para } \beta_2 = 0.$$

Para se verificar  $H_0: a = 0$  e  $d = 0$  (ausência de ambos efeitos), o modelo adequado é

$$y_j = \beta_0 + \varepsilon_j \text{ para } \beta_1 = 0 \text{ e } \beta_2 = 0.$$

Para a verificação da significância de cada hipótese, pode ser usado o teste de F ou t e, analogamente ao item anterior, ao se testar apenas um parâmetro ( $\beta_1 = 0$  ou  $\beta_2 = 0$ ), ter-se-á apenas 1 GL no numerador da estatística F, assim  $F = t^2$ .

## 2.3 Máxima verossimilhança usando a progênie F<sub>2</sub>

Quando indivíduos de duas linhagens endogâmicas são cruzados para produzir uma geração F<sub>1</sub>, pode-se observar que uma das linhagens parentais é homocigota para os alelos **M** do marcador e **Q** do QTL, ao passo que a outra é homocigota para os alelos **m** e **q**. Indivíduos da geração F<sub>1</sub>, que é heterocigota para ambos locos, são acasalados (ou autofertilizados) de modo a produzir a geração F<sub>2</sub>. Caso o QTL esteja ligado ao marcador, os indivíduos F<sub>2</sub> que possuírem diferentes genótipos deste marcador diferirão quanto aos valores da característica quantitativa (Weller, 1986). Se a ligação entre o marcador e o QTL é incompleta, então cada grupo de marcador conterá indivíduos com os três genótipos do QTL.

Nesse caso, a função de verossimilhança para os dados observados pode ser escrita como o produto das funções de densidade para cada indivíduo, sendo essas funções de densidade dadas pelo somatório do produto entre as funções de densidade marginais para cada genótipo do QTL ( $f(y_i | QQ)$ ,  $f(y_i | Qq)$ ,  $f(y_i | qq)$ ) e as probabilidades condicionais dos genótipos do QTL. Assim, tem-se que

$$L(\theta; y) \propto \prod \ell(y_i | M_i)$$

ou

$$L(\theta; y) \propto \prod_{i=1}^n \sum_{k=1}^N (f(y_i | Q_k) \cdot P(Q_k | M_i))$$

ou

$$L(\theta; y) \propto \prod_{i=1}^n \{ [f(y_i | QQ) \cdot P(QQ | M_i)] + [f(y_i | Qq) \cdot P(Qq | M_i)] + [f(y_i | qq) \cdot P(qq | M_i)] \}$$

em que

$\ell(y_i | M_i)$  = função de verossimilhança para um indivíduo com valor fenotípico  $y_i$  e genótipo do marcador  $M_i$ ;

$M_i$  = genótipo do marcador (MM, Mm ou mm) do <sup>i</sup>ésimo indivíduo;

$P(QQ | M_i)$ ,  $P(Qq | M_i)$  e  $P(qq | M_i)$  = probabilidades condicionais definidas anteriormente.

Como visto acima, a função de verossimilhança de cada indivíduo com valor fenotípico  $y_i$  e genótipo do marcador  $M_i$  ( $\ell(y_i | M_i)$ ) será

$$\ell(y_i | M_i) = \{ [f(y_i | QQ) \cdot P(QQ | M_i)] + [f(y_i | Qq) \cdot P(Qq | M_i)] + [f(y_i | qq) \cdot P(qq | M_i)] \}.$$

Todavia, nesse tipo de delineamento, os parâmetros a serem estimados são a frequência de recombinação  $r$ , as médias dos três genótipos QTL ( $\mu_{aa}$ ,  $\mu_{Aa}$ ,  $\mu_{aa}$ ), e as três variâncias, as quais normalmente são consideradas iguais.

As soluções ML para esses parâmetros, teoricamente, podem ser obtidas por meio de derivadas parciais de  $L(\theta; y) \propto \prod \ell(y_i | M_i)$ , ou do log desta função, em relação a cada um deles, igualando as derivadas parciais a zero e resolvendo as sete equações. No entanto, tais soluções analíticas são impraticáveis, devendo-se lançar mão de processos iterativos.

Os parâmetros do QTL podem ser estimados como funções das distribuições estatísticas dos genótipos do marcador. Usando-se somente a frequência de recombinação ( $r$ ), o primeiro e o segundo momento do QTL podem ser construídos segundo as equações apresentada por Weller (1986).

De acordo com Weller (1986), esperam-se maiores vieses na obtenção dos parâmetros referentes ao heterozigoto do que nos dos homozigotos, dado que estes últimos são estimados principalmente em função das médias e das variâncias da classe de QTL correspondente, ao passo que os heterozigotos são bastante influenciados por todos os genótipos QTL. Por essa razão, Weller (1986) sugeriu a combinação do método dos momentos e da ML para a obtenção das soluções dos parâmetros. Nessa combinação, inicialmente, são derivadas as estimativas para os parâmetros dos homozigotos pelo método dos momentos e, então, usa-se a ML para estimar a média e a variância do heterozigoto e a frequência de recombinação, reduzindo a um espaço paramétrico tri-dimensional. Como desvantagens, segundo o autor, este método pode produzir estimativas viesadas para QTLs de pequeno efeito ou com alta frequência de recombinação, além de não garantir que os resultados sejam de fato estimativas de ML.

## 2.4 Mapeamento por intervalo usando regressão

No mapeamento por intervalo, a obtenção de estimativas ML com o uso do algoritmo EM, segundo Haley e Knott (1992), pode ser complexa e computacionalmente demorada para atingir a convergência. Lynch e Walsh (1998) relataram, ainda, que a demanda computacional pode ser fator limitante na aplicabilidade de procedimentos de reamostragem, os quais requerem o cálculo de milhares de estimativas de ML por experimento.

A constatação de que os procedimentos que usam regressão simples fornecem estimativas muito aproximadas daquelas que utilizam a ML foi feita de maneira independente por Haley e Knott (1992), em progênies  $F_2$ , e por Martínez e Curnow (1992), em BC. Os resultados encontrados por Haley e Knott (1992) mostraram que a correlação entre os dois métodos sempre foi superior a 0,90, exceto em um caso, em que havia a presença de *outliers*.

Pelo fato de o método ser mais conhecido como **regressão de HALEY e KNOTT**, será abordado mais especificamente o trabalho destes autores, usando o delineamento F2. Todavia, em razão da facilidade de inclusão das fontes de variação em modelos de regressão, este método pode ser usado em outros delineamentos experimentais. Desse modo, as variações entre locais e famílias, por exemplo, podem ser consideradas desde que incluídas no modelo.

De modo análogo à estimação por ML, a regressão requer que a característica fenotípica seja normalmente distribuída, embora, este método possa ser usado na análise de dados categóricos.

Em seu trabalho, Harley e Knott (1992) desenvolveram um método de mapeamento por intervalo baseado na regressão múltipla do fenótipo sobre o tipo de marcador. As estimações da distância de mapa e do efeito são obtidas expressando-se o genótipo esperado de um suposto QTL como uma função dos genótipos de seus marcadores flanqueadores e sua frequência de recombinação.

Como pressuposições, admita-se que a variância dentro de cada classe de genótipos do QTL seja a mesma e tenha distribuição normal, e que os efeitos genotípicos dos três genótipos possíveis do QTL sejam iguais a

$$\mu_{QQ} = \mu + a, \quad \mu_{Qq} = \mu + d \quad \text{e} \quad \mu_{qq} = \mu - a,$$

em que  $\mu$  é a média dos homozigotos e  $a$  e  $d$  são os desvios aditivo e de dominância, respectivamente. As taxas de recombinação são iguais às consideradas no item anterior, ressaltando-se que neste método há a necessidade de se conhecer a frequência de recombinação entre os marcadores ( $r$ ). Para todas as análises, pressupõe-se ausência de interferência e o uso da função de mapeamento de Haldane para a conversão das distâncias para cM.

A média esperada para cada genótipo do marcador em termos do suposto QTL, ou seja, em termos da probabilidade condicional  $P(Q_i|M_i)$ , ou valores esperados de  $a$  e  $d$ , em termos de frequências de recombinação para cada um dos possíveis genótipos de uma população F<sub>2</sub>, podem ser encontrados em Haley e Knott (1992). Os valores esperados para outras gerações de segregação podem ser facilmente derivados de modo similar.

Expressando as probabilidades condicionais como um modelo de regressão linear múltipla, obtém-se

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i,$$

em que

$y_i$  = característica fenotípica;

$\beta_0$  = constante da regressão;

$\beta_1$  = diferença entre os genótipos homozigotos do QTL =  $a$ ;

$x_{1i}$  =  $P(QQ|M_i) - P(qq|M_i)$ ;

$\beta_2$  = efeito de dominância do QTL =  $d$ ;

$x_{2i}$  =  $P(Qq|M_i)$ ;

$\epsilon_i$  = variável aleatória  $\sim N(0, \sigma^2)$ .

Os valores esperados são funções das freqüências de recombinação entre QTL e cada um dos marcadores ( $r_1$  e  $r_2$ ) e a freqüência entre os marcadores ( $r$ ). Ajustando-se a regressão para valores crescentes (por exemplo, 0,1 cM ou 1cM) de  $r_1$ , a partir de zero, o qual corresponde à presença do QTL exatamente sobre o marcador  $M_1$ , obtém-se várias estimativas dos parâmetros  $\beta_0$ ,  $\beta_1$  e  $\beta_2$  (ou  $\mu$ ,  $a$  e  $d$ ). A posição em que for obtido o melhor ajuste, isto é, o maior coeficiente de determinação ( $R^2$ ) com o menor quadrado médio do resíduo, será a localização mais provável do QTL e a estimativa do parâmetro  $a$  ( $\hat{a}$ ) corresponderá ao efeito aditivo do QTL sobre a característica quantitativa, em que a significância de tal efeito é testada por meio do teste F.

Se a posição mais provável está em um intervalo com alta freqüência de recombinação, pode ser interessante refazer as análises utilizando outros marcadores mais próximos dentro do intervalo, o que possivelmente diminuirá a variância dentro do genótipo do marcador questionado.

A grande importância do trabalho de Haley e Knott (1992) não está somente na proposta de utilização da regressão, mas também na demonstração da similaridade dos resultados obtidos por regressão e por ML. Para tanto, esses autores propuseram um teste estatístico chamado **teste aproximado de razão de verossimilhança**, cuja expressão é

$$LR = n \cdot \log_e \left( \frac{SQR_{\text{modelo reduzido}}}{SQR_{\text{modelo completo}}} \right) = -n \cdot \log_e (1 - R^2),$$

o qual corresponde à razão entre a soma de quadrados do resíduo do modelo contendo o QTL ( $SQR_{\text{modelo completo}}$ ) e a soma de quadrados do resíduo sem o QTL ( $SQR_{\text{modelo reduzido}}$ ), em que  $n$  é o número de observações. O termo  $R^2$  refere-se ao coeficiente de determinação (ou percentagem da variância explicada pelo modelo), somente podendo ser usado se outros efeitos fixos não forem incluídos na análise.

As críticas normalmente feitas a este método referem-se ao tempo gasto na confecção das análises e à sua fragilidade para indicar a posição do QTL quando a suposição de normalidade é falha. Também, como ela analisa um intervalo por vez, os efeitos de possíveis QTLs ligados em regiões flaqueadoras são ignorados, e as estimativas obtidas serão viesadas dentro do intervalo considerado, aumentando a possibilidade de indicar erroneamente a existência de um QTL entre eles (*ghost* QTL).

Para ilustrar este método, será usado o exemplo apresentado por Lynch e Walsh (1998) no Cap. 15, páginas 455 e 456. Nesse exemplo, os autores analisaram dados hipotéticos provenientes de um delineamento  $F_2$ , considerando dois marcadores flanqueadores ( $M_1$  e  $M_2$ ) separados por uma frequência de recombinação ( $r$ ) igual a 0,30 e ausência de interferência.

Os genótipos dos marcadores, bem como os valores da característica fenotípica ( $Y$ ) e o cálculo dos valores esperados para o efeito aditivo ( $a$ ) e de dominância ( $d$ ), são apresentados no seguinte programa do SAS:

/\*Delineamento F2 usando a regressão de Haley e Knott\*/

OPTIONS LS = 78 PS = 64;

DATA UM;

INPUT INDIVID 1-2 GENOTIPO \$ 4-11 Y 13-15 .1;

CARDS;

1 M1n1M2n2 3.9

2 M1M1M2M2 5.6

3 M1n1M2M2 3.7

4 n1n1M2n2 3.9

5 M1M1M2n2 5.3

6 n1n1n2n2 1.1

7 M1n1M2M2 3.6

8 M1M1M2M2 5.4

9 M1n1M2M2 3.7

10 M1n1M2n2 3.3

;

DATA DOIS; SET UM;

r1 = 0; /\* Frequencia de recombinação entre QTL e marcador 1\*/

r = 0.3; /\* Frequencia de recombinação entre marcador1 e marcador 2\*/

r2 = (0.3-r1)/(1-2\*r1); /\* Frequencia de recombinação entre QTL e marcador 2 (com i=0)\*/

r1q = r1\*r1;

r2q = r2\*r2;

rq = r\*r;

r11 = (1-r1)\*(1-r1);

r22 = (1-r2)\*(1-r2);

r12 = (1-r)\*(1-r);

X1=0;

X2=0;

```

/* Definição dos valores esperados de 'a' e 'd' */
IF GENOTIPO EQ 'M1M1M2M2' THEN X1 = ((r11*r22) - (r1q*r2q))/(r12);
IF GENOTIPO EQ 'M1M1M2M2' THEN X2 = (2*r1*(1-r1)*r2*(1-r2))/(r12);

IF GENOTIPO EQ 'M1M1M2n2' THEN X1 = (((r11)*r2*(1-r2))-(r1q*r2*(1-r2)))/(r*(1-r));
IF GENOTIPO EQ 'M1M1M2n2' THEN X2 = ((r1*(1-r1)*r22) + (r1*(1-r1)*r2q))/(r*(1-r));

IF GENOTIPO EQ 'M1M1n2n2' THEN X1 = ((r11*r2q) - (r1q*(r22)))/(rq);
IF GENOTIPO EQ 'M1M1n2n2' THEN X2 = (2*r1*(1-r1)*r2*(1-r2))/(rq);

IF GENOTIPO EQ 'M1n1M2M2' THEN X1 = ((r1*(1-r1)*r22) - (r1*(1-r1)*r2q))/(r*(1-r));
IF GENOTIPO EQ 'M1n1M2M2' THEN X2 = (((r11)*r2*(1-r2)) + (r1q*r2*(1-r2)))/(r*(1-r));

IF GENOTIPO EQ 'M1n1M2n2' THEN X1 = 0;
IF GENOTIPO EQ 'M1n1M2n2' THEN
X2 = (r1q*r2q) + (r1q*r22) + (r11*r2q) + (r11*r22))/(rq + r12);

IF GENOTIPO EQ 'M1n1n2n2' THEN X1 = ((r1*(1-r1)*r2q) - (r1*(1-r1)*r22))/(r*(1-r));
IF GENOTIPO EQ 'M1n1n2n2' THEN X2 = (((r11)*r2*(1-r2)) + (r1q*r2*(1-r2)))/(r*(1-r));

IF GENOTIPO EQ 'n1n1M2M2' THEN X1 = ((r1q*r22) - (r1*(1-r1)*r2q))/(rq);
IF GENOTIPO EQ 'n1n1M2M2' THEN X2 = (2*r1*(1-r1)*r2*(1-r2))/(rq);

IF GENOTIPO EQ 'n1n1M2n2' THEN X1 = ((r1q*r2*(1-r2)) - (r11*r2*(1-r2)))/(r*(1-r));
IF GENOTIPO EQ 'n1n1M2n2' THEN X2 = ((r1*(1-r1)*r22) + (r1*(1-r1)*r2q))/(r*(1-r));

IF GENOTIPO EQ 'n1n1n2n2' THEN X1 = ((r1q*r2q) - (r11*r22))/(r12);
IF GENOTIPO EQ 'n1n1n2n2' THEN X2 = (2*r1*(1-r1)*r2*(1-r2))/(r12);

DATA TRES; SET DOIS;
TITLE 'Análise de regressão para r1 = 0';

PROC REG;
MODEL Y = X1 X2;
RUN;

```

é

**Análise de regressão para  $r_1 = 0$**

Model: MODEL1

Dependent Variable: Y

**Analysis of Variance**

Source	DF	Sum of Squares	Mean Square	F Value	Prob > F
Model	2	11.28633	5.64317	9.499	0.0101
Error	7	4.15867	0.59410		
C Total	9	15.44500			

Root MSE	0.77078	<b>R-square</b>	<b>0.7307</b>
Dep Mean	3.95000	Adj R-sq	0.6538
C.V.	19.51331		

**Parameter Estimates**

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob >  T
INTERCEP( $\mu$ )	1	3.966667	0.35180938	11.275	0.0001
X1 (a)	1	1.466667	0.35180938	4.169	0.0042
X2 (d)	1	-0.326667	0.49253313	-0.663	0.5284

Os resultados das análises considerando diferentes valores de  $r_1$  são apresentados na Tabela 2 e na Figura 1.

Tabela 2. Valores da média e dos efeitos aditivo e de dominância, em análises de regressão de Haley e Knott, considerando diferentes freqüências de recombinação entre QTL e o marcador  $M_1$

Freqüência de recombinação ( $r_1$ )	$\hat{\mu}$	$\hat{a}$	$\hat{d}$	$R^2$	LOD <sup>1</sup>
0,00	3,967	1,467	-0,327	0,731	2,846
0,03	3,956	1,563	-0,378	0,745	2,961
0,06	3,937	1,661	-0,428	0,756	3,056
0,09	3,903	1,757	-0,459	0,760	3,098
0,12	3,833	1,839	-0,426	0,754	3,041
0,15	3,698	1,895	-0,257	0,732	2,859
0,18	3,469	1,916	0,123	0,698	2,599
0,21	3,171	1,908	0,670	0,665	2,371
0,24	2,903	1,876	1,164	0,643	2,235
0,27	2,757	1,800	1,399	0,626	2,133
0,30	2,750	1,650	1,350	0,597	1,973

$${}^1\text{LOD} = -n \cdot \log_e(1 - R^2) / 4,61$$

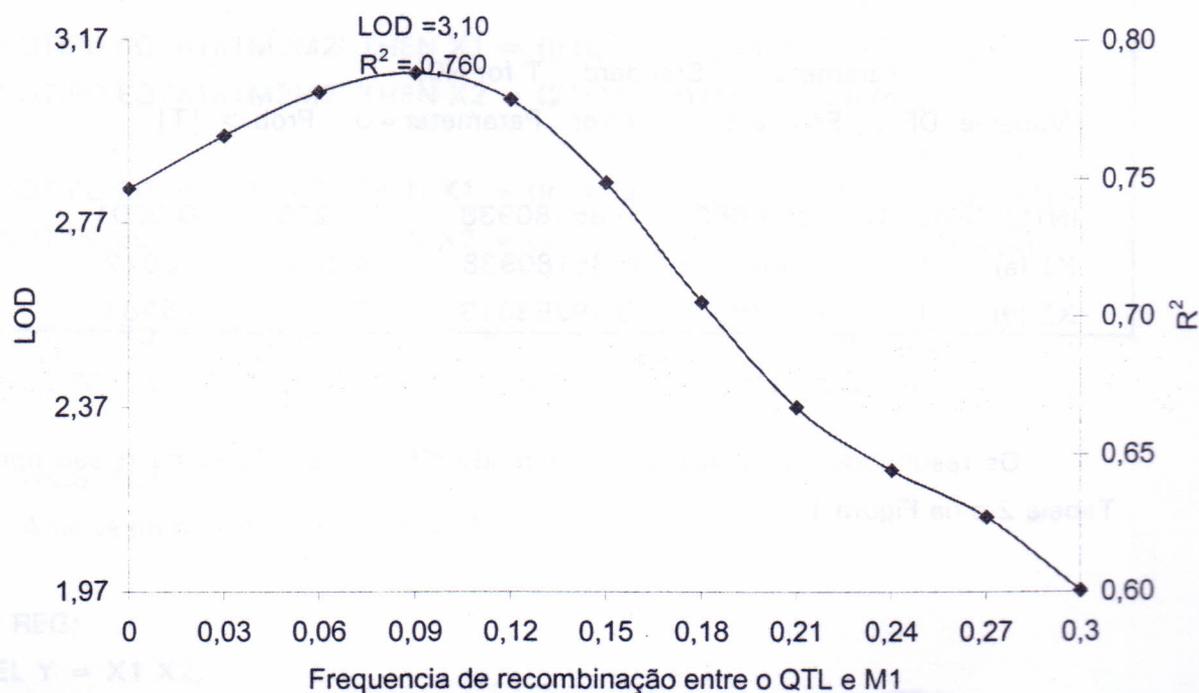


Figura 1. Gráfico do coeficiente de determinação ( $R^2$ ) das análises de regressão, em função de diferentes freqüências de recombinação.

De acordo com os resultados obtidos, o valor máximo do  $R^2$  ocorre quando  $r_1$  é igual a 0,09, e as estimativas dos parâmetros são  $\hat{\mu} = 3,90$ ,  $\hat{a} = 1,76$  e  $\hat{d} = -0,46$ , sugerindo que um suposto QTL está presente neste intervalo, distante  $r_1 = 0,09$  do loco do marcador  $M_1$ , com médias genotípicas estimadas iguais a

$$\hat{\mu}_{QQ} = \hat{\mu} + \hat{a} = 5,66, \quad \hat{\mu}_{Qq} = \hat{\mu} + \hat{d} = 3,44 \quad e \quad \hat{\mu}_{qq} = \hat{\mu} - \hat{a} = 2,14.$$

O lod score ("*log of the odds ratio score*"), proposto por Morton (1955), é outro teste bastante usado na detecção de QTLs. Para determinado valor da taxa de recombinação ( $r$ ) obtido na análise de máxima verossimilhança, o lod score (LOD) se relaciona ao LR por

$$LOD = \log_{10} \left[ \frac{\max L(r)}{\max L(\theta; y)} \right] = \frac{LR_r}{2 \ln 10} \cong \frac{LR_r}{4,61}$$

mostrando que o LOD é simplesmente o  $\log_{10}$  LR, ou então, o produto entre uma constante ( $1/4,61$ ) e o LR. Se esta proporção for maior do que determinado valor limiar (*threshold value*), a hipótese de nulidade será rejeitada e o QTL e o marcador serão considerados ligados. Em testes estatísticos clássicos, a decisão de rejeição da hipótese de nulidade depende diretamente da escolha da probabilidade  $\alpha$  (ou nível de significância) de se cometer o erro tipo I, isto é, rejeitar  $H_0$  quando ela é verdadeira. O uso do valor limiar para o LOD não pode ser justificado do mesmo modo, pois ele não está ligado a um valor fixo de  $\alpha$ . Normalmente, o valor limiar adotado nos trabalhos é igual a 3, como sugerido por Morton (1955), sendo esse valor obtido por meio das relações entre as probabilidades de ocorrência de 0,1% para o erro tipo I ( $A \approx 1000$ ) e 1% para o erro tipo II ( $B \approx 0,01$ ), estabelecidas por Wald (1947), citado por MORTON (1955). Assim, o  $\log_{10}(A) = 3 = 10^3$  significa que a hipótese alternativa é 1000 vezes mais provável do que a de nulidade. Todavia, nas análises de dados provenientes de delineamentos experimentais, dependendo do tipo de população analisada, diferentes valores para o limiar podem ser utilizados com o mesmo nível de significância, pois a própria estrutura do delineamento pode aumentar ou diminuir tal nível.

Para verificar a evidência significativa da presença do suposto QTL, pode-se converter o valor do LOD para LR, multiplicando-se por 4,61 e comparando o valor obtido ao valor tabelado de  $\chi^2$  com três graus de liberdade (para  $a$ ,  $d$  e  $r_1$ ). Assim, tem-se que

$$Pr[\chi_3^2 > -n \cdot \ln(1 - r^2)] = Pr[\chi_3^2 > (3,10) \cdot (4,61)] = Pr[\chi_3^2 > 14,02] = 0,003,$$

mostrando que o efeito do QTL é significativo.

### 3. Aplicação de modelos lineares mistos

A predição do valor genético de um animal, por meio das equações de modelos mistos (EMM), é baseada no parentesco genético aditivo entre os indivíduos. Na construção da matriz de parentesco, normalmente, são utilizadas as probabilidades de que dois indivíduos possuam, em determinado loco, genes idênticos por descendência (IBD), o que significa dizer que os genes que os dois animais possuem são cópias exatas oriundas de um ancestral comum. Desta forma, nenhum conhecimento sobre a real semelhança entre dois indivíduos é usado. Fernando e Grossman (1989) e Van Arendonk et al. (1994) mostraram que a informação a respeito de um único marcador pode ser usada, em um modelo animal, para ajustar os efeitos aditivos para QTLs ligados e demais efeitos poligênicos aditivos, na análise de determinada característica quantitativa. Goddard (1992) estendeu o modelo proposto por Fernando e Grossman (1989), de modo a possibilitar a inclusão de informações de mais de um marcador. Posteriormente, Bink et al. (1998) estabeleceu uma metodologia, baseada na amostragem de Gibbs, que permite recuperar informações de animais que não foram genotipados.

Meuwissen e Goddard (1996), citados por Bink et al. (1998), apresentaram um método no qual a matriz de co-variâncias dos efeitos do QTL ligado ao marcador é aproximada. Tal aproximação reduz os requerimentos computacionais e facilita a inclusão de animais sem informação sobre o genótipo do marcador.

Para situações em que existam informações sobre um marcador em apenas duas gerações, somente a variância dos efeitos de substituição dos alelos pode ser estimada, enquanto que no caso de marcadores flanqueadores, a posição e a variância do QTL podem ser estimadas.

Com registros dos fenótipos e dos genótipos dos marcadores colhidos em várias gerações, torna-se possível separar a posição e o efeito do QTL e estimar a variância e a distância entre QTL e marcador. Além disso, estimativas das variâncias dos efeitos poligênicos e residual também podem ser obtidas.

Basicamente, os métodos acima diferem em relação à obtenção da matriz de (co)variâncias genótípicas (também chamada de matriz de parentesco gamético) e, ou, sua inversa. Em cada um deles são obtidos preditores para o efeito genético aditivo relacionado ao QTL ligado ao marcador e para o efeito genético aditivo de outros QTLs não ligados. Na metodologia proposta por Van Arendonk et al. (1994), esses autores utilizaram um método para construção da matriz de parentesco e de sua inversa, por meio da teoria de partição de matrizes (Tier e Sölkner, 1993), de forma a incorporar a informação sobre um locus de marcador na predição dos valores genéticos. Além disso, os autores mostraram como utilizar a informação proveniente de múltiplos marcadores na predição de somente um efeito aleatório para cada indivíduo. Ainda, esta metodologia permite incluir informações sobre marcadores

flanqueadores, vários marcadores ligados ao mesmo QTL ou mesmo para mais de um QTL, além de incorporar animais sem informação sobre o marcador.

A despeito da existência das metodologias, aparentemente não existe ainda algoritmo computacional que possa ser usado com rapidez e eficiência na análise de dados moleculares.

#### 4. Aplicação de modelos aleatórios

De acordo com Amos e Elston (1989), métodos robustos, baseados em modelos aleatórios, têm sido desenvolvidos (Haseman e Elston, 1972; 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 co-variância) entre dois indivíduos aparentados. A co-variância 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. Supõe-se, portanto, que ele seja a média da proporção de alelos IBD compartilhados por dois indivíduos, ou, em outras palavras, é igual ao numerador do coeficiente de parentesco de Wright, isto é,  $1/2$  para irmãos completos e  $1/4$  para meios-irmãos.

De acordo com Martinez et al. (1999), para o mesmo tipo de parentesco, as proporções de 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 o mesmo alelo do QTL são geneticamente mais semelhantes do que outros dois irmãos que não compartilham nenhum. Obviamente, como não se pode observar diretamente os alelos do QTL, essa proporção é tomada a partir dos genótipos observados dos marcadores.

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, em que há predominância de famílias pequenas de irmãos completos. Esse método é baseado na regressão linear dos quadrados das diferenças fenotípicas entre dois irmãos, dentro da família, em função da proporção de IBD compartilhados por eles. Apesar de ser considerado robusto para diversas distribuições dos dados e apesar da 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, em que 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-Soressen e Roberson, 1961), para o caso de grandes famílias de meias-irmãs.

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 procedimento de pares de irmãos usando o mapeamento por intervalo. Entretanto, apesar do maior poder estatístico, esse método é baseado em quadrados mínimos, não otimizando o uso de toda a informação contida nos dados, como as propriedades distribucionais, conforme ocorre com a máxima verossimilhança (ML).

Segundo Martinez e Vukasinovic (2000), o primeiro método para estimar a variância genética devida à determinada região do cromossomo, com base na ML, chamado de 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 de ambiente comum partilhado por dois irmãos. Embora esses métodos utilizassem 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 baseado no método de ML desenvolvido por Goldgar (1990), no qual se considera a distribuição normal dos valores genotípicos do QTL e o ajustamento do QTL e do efeito poligênico como efeitos aleatórios para a análise de famílias de irmãos completos. Martinez et al. (1999) estenderam o procedimento para situações em que há prevalência de famílias de meias-irmãs, como em gado de leite, e mostraram que QTL com efeitos relativamente grandes podem ser detectados com alto poder e localizados com bastante acurácia, especialmente se grande número de famílias e marcadores polimórficos são usados.

A seguir, serão descritas as principais metodologias que utilizam modelos aleatórios.

#### **4.1 Metodologia proposta por Goldgar (1990)**

Apesar do maior poder estatístico, o método proposto por Fulker e Cardon (1994) é baseado em quadrados mínimos, o que não otimiza o uso de toda a informação contida nos dados, como as propriedades distribucionais, ao contrário da máxima verossimilhança (ML), que a otimiza.

O primeiro método para estimar a variância genética devida à determinada região no cromossomo, com base na ML, chamado método de pontos múltiplos de IBD, foi desenvolvido por Goldgar (1990), e é baseado na estimativa da proporção esperada de material genético em um cromossomo particular (ou braço) compartilhado por cada indivíduo de dado par de irmãos, o qual possui vários locos de marcadores. Estas estimativas são usadas para particionar estatisticamente a variância genética em efeitos dos locos dos marcadores de uma região específica do cromossomo.

Tal metodologia, segundo Goldgar (1990), permite detectar três tipos de variância genética: (1) variação devida ao loco de efeito maior; (2) variação devida a vários locos presentes na mesma região ou cromossomo; e (3) variação devida a vários locos não-ligados, cada um tendo influência moderada sobre a característica.

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

$$y_{ij} = \mu + g_{ij} + a_{ij} + e_{1j},$$

em que

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

$\mu$  = média geral;

$g_{ij}$  = efeito genético aditivo do QTL  $\sim N(0, \sigma_g^2)$ ;

$a_{ij}$  = efeito poligênico  $\sim N(0, \sigma_a^2)$ ;

$e_{ij}$  = efeito de ambiente  $\sim N(0, \sigma_e^2)$ .

Todos os efeitos aleatórios do modelo seguem distribuição normal, entretanto, se  $\sigma_a^2$  e  $\sigma_e^2$  são suficientemente grandes para tornar os dados normalmente distribuídos, não é exigida a distribuição normal do efeito do QTL.

Para o uso desse modelo, pressupõe-se ausência de interferência e de dominância, os crossing-overs seguem a distribuição de Poisson e ocorrem a uma distância de  $\lambda$  Morgans, e são uniforme e independentemente distribuídos na região analisada.

A variância de  $y_{ij}$ , considerando-se equilíbrio de ligação, é

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

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

$$\text{Cov}(y_{ij}, y_{i'j'}) = \pi_q \sigma_g^2 + \frac{1}{2} \sigma_a^2 \quad (\text{irmãos completos})$$

ou

$$\text{Cov}(y_{ij}, y_{i'j'}) = \pi_q \sigma_g^2 + \frac{1}{4} \sigma_a^2 \quad (\text{meios-irmãos}),$$

em que  $\pi_q$  é a proporção de alelos IBD do suposto QTL compartilhado pelos irmãos.

Os coeficientes associados à variância poligênica podem ser  $\frac{1}{2}$  ou  $\frac{1}{4}$  porque, em média, espera-se que dois irmãos completos ou dois meios-irmãos compartilhem, respectivamente, tais proporções de IBD. As proporções de IBD do QTL ( $\pi_q$ ) serão diferentes para cada par de irmãos, podendo variar de 0 a 1, para irmãos completos, ou de 0 a  $\frac{1}{2}$ , para meios-irmãos.

Em qualquer estrutura de família, para a estimação dos componentes de variância, deve-se substituir  $\pi_q$  por seu valor estimado,  $\hat{\pi}_q$ . A co-variância entre os irmãos  $j$  e  $j'$  dentro da família  $i$  é

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

em que

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

e

$$r_i = \hat{\pi}_q h_g^2 + \frac{1}{2} h_a^2 \text{ (para irmãos completos)}$$

ou

$$r_i = \hat{\pi}_q h_g^2 + \frac{1}{4} h_a^2 \text{ (para meios-irmãos)}.$$

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

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, a função de densidade conjunta das observações dentro das famílias é:

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

em que  $y_i = [y_{i1}, y_{i2}, y_{i3}, \dots, y_{ik}]'$  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.

O log da verossimilhança para  $n$  famílias independentes é

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

A função de verossimilhança relaciona a posição do QTL flanqueado pelos dois marcadores por meio de  $r$ . Os parâmetros desconhecidos a serem estimados são  $\mu$ ,  $\sigma^2$ ,  $h_g^2$ ,  $h_a^2$  e  $\theta_{1q}$ . Para maximizar  $L$  no mapeamento por intervalo, considera-se a fração de recombinação entre o primeiro marcador e o suposto QTL ( $\theta_{1q}$ ) como constante e, gradualmente, vai-se aumentando seu valor e, conseqüentemente, diminuindo-se a distância entre o QTL e o marcador posicionado do lado direito ( $\theta_{2q}$ ), 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  $\theta_{1q}$ , que maximiza  $L$  e a hipótese nula  $h_g^2 = 0$  pode ser testada por meio do teste de razão de verossimilhança.

Segundo GOLDGAR (1990), o método apresenta como vantagens a consideração da estrutura de *pedigree* e a possibilidade de se trabalhar com dados genotípicos multilocos, apresentando maior poder de detecção e evitando o problema estatístico de analisar famílias como um conjunto de pares de irmãos mutuamente independentes.

#### 4.2 Metodologia proposta por Xu e Atchley (1995)

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 variação significativa devida 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, também, que a probabilidade de se encontrar dois QTLs no mesmo intervalo seja reduzida, podendo ser negligenciada. Assim, esse método foi desenvolvido 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 QTLs. Para ilustrar o método, será

descrito somente o modelo com múltiplos QTLs, dado que o modelo com um QTL é semelhante ao de Goldgar (1990).

Caso haja múltiplos QTLs no mesmo grupo de ligação, a estimação tende a ser viesada, em virtude da interferência causada pelos QTLs localizados no mesmo cromossomo, mas fora do intervalo testado (Jansen, 1993; Jansen, 1994; Zeng, 1993, Zeng, 1994). Desse modo, o modelo para múltiplos QTLs pode ser descrito como:

$$y_{ij} = \mu + a_{ij} + \sum_{k=1}^S u_{ij}^k + g_{ij} + \sum_{r=1}^W u_{ij}^r + e_{1j},$$

em que

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

$\mu$  = média geral;

$a_{ij}$  = efeito poligênico  $\sim N(0, \sigma_a^2)$ ;

$u_{ij}^k$  = efeito dos QTL localizados à esquerda do suposto QTL analisado, em que  $S$  é o número de locos;

$g_{ij}$  = efeito genético aditivo do QTL  $\sim N(0, \sigma_g^2)$ ;

$u_{ij}^r$  = efeito dos QTL localizados à direita do suposto QTL analisado, em que  $W$  é o número de locos;

$e_{ij}$  = efeito de ambiente  $\sim N(0, \sigma_e^2)$ .

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

$$\text{Var}(y_{ij}) = \sigma^2 = \sigma_a^2 + \sum_{k=1}^S \sigma_k^2 + \sigma_g^2 + \sum_{r=1}^W \sigma_r^2 + \sigma_e^2$$

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

$$\text{Cov}(y_{ij}, y_{i'j'}) = \frac{1}{2}\sigma_a^2 + \sum_{k=1}^S \pi_{ik}\sigma_k^2 + \pi_{iq}\sigma_g^2 + \sum_{r=1}^W \pi_{ir}\sigma_r^2.$$

Deve-se considerar que os  $\pi$ 's são as proporções IBD dos QTLs e elas não são observáveis. Xu e Atchley (1995) definiram, teoricamente, a co-variância condicional dos  $\pi$ '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 QTLs. Todavia, a técnica não se mostrou eficiente em função do número de parâmetros a ser estimado.

Usando as mesmas propriedades das variáveis indicadoras, Zeng (1993) propôs outro método, em que se pressupõe que um marcador seja suficiente para absorver a correlação entre o loco QTL à sua esquerda e 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  $\pi_{iL}$  e  $\pi_{iR}$  como os valores IBD dos marcadores à esquerda (**L**) e à direita (**R**), respectivamente, e  $\theta_{kL}$  e  $\theta_{rR}$  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 co-variância condicional entre irmãos completos é

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

em que

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

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

$$h_k^2 = \frac{\sigma_k^2}{\sigma^2}$$

$$h_r^2 = \frac{\sigma_r^2}{\sigma^2}.$$

Para 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 ML sob a hipótese nula ( $L_0$ ) é usada para a obtenção do teste de razão de verossimilhança, que segue distribuição de qui-quadrado com  $2 \geq GL > 1$  sob  $H_0$ . Um GL é devido ao ajustamento para  $h_g^2$  e o outro para o ajustamento da posição do QTL, dependendo do tamanho do intervalo. O outro valor pode ser menor do que 2 (dois) 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  $H_0$ .

## 5. 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 estimation with incomplete marker data. *Gen. Sel. Evol.*, 30:45-58.
- BOILARD, M., BOICHARD, D. 2002. Use of maternal information for QTL detection in a (grand)daughter design. *Genet. Sel. Evol.*, 34:335-352.
- FALCONER, D.S., MACKAY, T.F.C. 1996. Introduction to quantitative genetics. 4<sup>th</sup> 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.
- 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.
- HALEY, C.S. , KNOTT, S.A. 1992. A simple regression method for ampping 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.
- JANSEN, R.C. 1993. Interval mapping of multiple quantitative trait loci. *Genetics*, 135:205-211.
- JANSEN, R.C. 1994. Controlling the Type I and Type II errors in mapping quantitative trait loci. *Genetics*, 138: 871-881.
- 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.
- LIU, B.-H. 1998. Statical genomics: linkage, mapping, and QTL analysis. 1<sup>st</sup> ed. , Ed. CRC Press, Boca Raton, FL.

- LYNCH, M., WALSH, B. 1998. Genetics and analysis of quantitative traits. 1<sup>st</sup> ed. , Ed. Sinauer Associates, Inc., Sunderland, MA.
- MARTÍNEZ, O., CURNOW, R.N. 1992. Estimating the locations and the sizes of the effects of quantitative trait loci using flanking markers. *Theor. Appl. Genet.*, 85:480-488.
- 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 meios-irmãos. *Rev. Bras. Zoot.*, 29:443-451.
- MARTINEZ, M. L. 2001. Métodos estatísticos para a detecção de locos de características quantitativas. In: REGITANO, L.C.A. & COUTINHO L.H.(Ed.) *Biologia molecular aplicada à produção animal*. 1.ed. Brasília: Embrapa Informação Tecnológica, 2001, p. 79-102.
- MORTON, N.E. 1955. Sequential tests for the detection of linkage. *Am. J. Hum. Genet.*, 10:344-349.
- NELDER, J.A., MEAD, R. 1965. A simplex method for function minimization. *Comput. J.*, 7:308-313.
- NEIMANN-SORENSEN, A., ROBERTSON, A. 1960. 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 Insitute Inc, Cary, NC, 4<sup>th</sup> 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.
- TIER, B., SÖLKNER, J. 1993. Analysing gametic variation with an animal model. *Theor. Appl. Genet.*, 85:868-872.
- 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.
- XU, S., ATCHLEY, W.R. 1995. A random model approach to interval mapping of quantitative trait loci. *Genetics*, 141: 1189-1197.
- WELLER, J.I. 1986. Maximum likelihood techniques for the mapping and analysis of quantitative trait loci with the aid of genetic markers. *Biometrics*, 42: 627-640.
- 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.
- ZENG, Z.-B. 1994. Precision mapping of quantitative trait loci. *Genetics*, 136: 1457-1468.

Realização:

**Embrapa**

*Pecuária Sudeste*

Patrocinadores: **BIOSYSTEMS**

**Amersham Bioscience**

Apoio:



**Embrapa**

---

**Empresa Brasileira de Pesquisa Agropecuária**  
**Embrapa Pecuária Sudeste**  
**Rod. Washington Luiz, km 234 -Fazenda Canchim**  
**C.Postal, 339 - São Carlos,SP - 13560-970**  
**Tel.: (16) 261 5611 - Fax: (16) 261 5754**  
**Email: [sac@cppse.embrapa.br](mailto:sac@cppse.embrapa.br)**  
**Home Page: [www.cppse.embrapa.br](http://www.cppse.embrapa.br)**