

ADMILSON DA COSTA E SILVA

**MAPEAMENTO DE MARCAS MOLECULARES E IDENTIFICAÇÃO DE
QTL COM BASE EM DESEQUILÍBRIO DE FASE GAMÉTICA**

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

**VIÇOSA
MINAS GERAIS – BRASIL
2008**

ADMILSON DA COSTA E SILVA

**MAPEAMENTO DE MARCAS MOLECULARES E IDENTIFICAÇÃO DE
QTL COM BASE EM DESEQUILÍBRIO DE FASE GAMÉTICA**

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

APROVADA: 26 de março de 2008.

Prof. Cosme Damião Cruz
(Co-orientador)

Prof. Fabyano Fonseca e Silva
(Co-orientador)

Dra. Cláudia Teixeira Guimarães

Dr. Jurandir Vieira de Magalhães

Prof. José Marcelo Soriano Viana
(Orientador)

*"Tudo tem seu tempo, há um momento oportuno para cada empreendimento
debaixo do céu. Tempo para plantar e tempo de colher a planta."*

Ecl. 3-1.2

Aos meus pais, Amador e Maria de Lourdes.

Aos meus irmãos, Luciano, Álisson e Leidiane.

À minha esposa, Vanessa, à nossa filha Vitória.

DEDICO

AGRADECIMENTOS

A Deus, pelas oportunidades que tenho recebido.

Aos meus pais, irmãos e demais familiares, pelo apoio aos meus estudos.

À Universidade Federal de Viçosa (UFV), pela oportunidade de realizar este curso.

Ao professor José Marcelo Soriano Viana, pela orientação, pelo apoio e pela confiança, pessoa exemplar em caráter e dedicação.

Ao professor e co-orientador Cosme Damião Cruz, pelas orientações, pela amizade e dedicação ao ensino e à pesquisa.

Ao professor e co-orientador Fabyano Fonseca e Silva, pelos ensinamentos, pela amizade e pelas sugestões sobre a dissertação.

À Dra. Cláudia Teixeira Guimarães e ao Dr. Jurandir Vieira de Magalhães, pela disposição em participar da banca e pelas considerações feitas.

Ao companheiro Vinícius Ribeiro Faria, pela colaboração e amizade.

Aos colegas Leonardo Bhering, Pedro Ivo e Thiago, pela colaboração e amizade.

Aos colegas Ahmed, Fábio, Isis, Ramon, Keny, Magno, Gabriel e Rodrigo, pela amizade.

Aos companheiros das repúblicas “Vaca Magra” e “Retiro dos Peões”: Gilberto (Bacana), Odirley (Batata), Fabrício, Leonardo (Violeiro) e Everton (Dudim).

Aos colegas do Grupo de Estudos em Genética e Melhoramento-GenMelhor, pelo convívio e pela oportunidade de fazer parte deste grupo.

À EMATER-MG pela licença de trabalho concedida.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pelo auxílio financeiro, indispensável à condução deste trabalho.

A todos os profissionais e colegas do programa de Pós-Graduação em Genética e Melhoramento da Universidade Federal de Viçosa.

A todos os profissionais em educação que contribuíram para minha formação acadêmica e também ao povo brasileiro, que por meio de impostos financia o ensino e a pesquisa pública.

Enfim, a todos que colaboraram para esta conquista, meus sinceros agradecimentos.

BIOGRAFIA

ADMILSON DA COSTA E SILVA, filho de Amador Francisco da Silva e de Maria de Lourdes da Costa e Silva, nasceu em 25 de agosto de 1973, em Pitangui, Minas Gerais.

Cursou o ensino fundamental na Escola Estadual Cel. Pedro Lino, na Escola Municipal Dalila Vieira e na Escola Estadual Dr. José Gonçalves, em Martinho Campos - MG.

Em janeiro de 1997, concluiu o curso profissionalizante de Técnico em Agropecuária na Central de Ensino e Desenvolvimento Agrário de Florestal (CEDAF), em Florestal - MG.

De janeiro a novembro de 1997, estagiou na Companhia Vale do Rio Doce (CVRD), Departamento de Apoio Operacional Civil-Meio Ambiente, em Itabira - MG.

Em dezembro de 1997, foi contratado pela Empresa de Assistência Técnica e Extensão Rural de Minas Gerais (EMATER - MG) como Extensionista Agropecuário I.

Em 2000, ingressou no curso de Ciências Biológicas do Centro Universitário de Patos de Minas (UNIPAM), concluindo a Licenciatura em dezembro de 2003 e o Bacharelado em dezembro de 2004.

Em maio de 2006, iniciou o Programa de Pós-Graduação em Genética e Melhoramento, em nível de Mestrado na Universidade Federal de Viçosa, Viçosa-MG, submetendo-se à defesa da dissertação em 26 de março de 2008.

SUMÁRIO

RESUMO	x
ABSTRACT.....	xiii
1- INTRODUÇÃO.....	1
2- REVISÃO DE LITERATURA.....	3
2.1- Desequilíbrio de Ligação.....	3
2.1.1- Desequilíbrio de Ligação em Genômica de Plantas	6
2.2- Marcadores Moleculares	7
2.3- Mapas Genéticos	11
2.4- Populações de Mapeamento de QTL.....	12
2.5- Detecção de QTL	14
2.6- Métodos de Mapeamento.....	16
3- MATERIAL E MÉTODOS	18
3.1- Aspectos teóricos.....	18
3.1.1- Análise de Desequilíbrio de Fase Gamética	18
3.1.2- Identificação de QTL pela Análise de Marcas Simples.....	23
3.2- Simulação dos Dados	29
3.2.1- Mapeamento de Marcas Moleculares.....	29

3.2.2- Detecção de QTL.....	33
4- RESULTADOS E DISCUSSÃO	36
4.1- Mapeamento de Marcas Moleculares	36
4.2- Detecção de QTL	40
4.2.1- Avaliação do Poder de Detecção do QTL.....	40
4.2.2- Teste de Dominância no Loco do Marcador	45
4.2.3- Avaliação dos Efeitos de Substituição de marca e dos Desvios de Dominância no Loco do Marcador	49
5- CONCLUSÕES.....	57
6- REFERÊNCIAS BIBLIOGRÁFICAS	59

RESUMO

SILVA, Admilson da Costa e, M.Sc., Universidade Federal de Viçosa, março de 2008. **Mapeamento de marcas moleculares e identificação de QTL com base em desequilíbrio de fase gamética.** Orientador: José Marcelo Soriano Viana. Co-orientadores: Cosme Damião Cruz e Fabyano Fonseca e Silva.

No estudo de ligação entre dois locos marcadores e análise de QTL é preciso realizar cruzamentos apropriados para o mapeamento das populações. No entanto, a realização de cruzamentos pode ser difícil em algumas espécies. Nestes casos, tem-se utilizado desequilíbrio de ligação (LD) ou de fase gamética. Este trabalho, realizado por meio de simulação de dados, teve como objetivos avaliar a eficiência do mapeamento de marcas moleculares e identificar o QTL em populações não-endogâmicas e não-estruturadas em famílias, com base em desequilíbrio de fase gamética. Para o mapeamento de marcas moleculares foram simulados genomas com nível de saturação alto (distância menor que 5 cM) e aleatório (distância entre 0 e 20 cM), cinco grupos de ligação e 20 marcas por grupo. Para obtenção dos genomas amostrais foram simulados compostos de segunda geração, com 1.000 e 200 indivíduos, derivados do cruzamento entre duas populações parentais, com tamanho 200, sob quatro níveis (1,0, 0,9, 0,8 e 0,7) de diferença mínima de frequência de mesma marca entre os genitores. A partir dos genomas amostrais foi avaliada

a eficiência de recuperação dos genomas paramétricos. Na detecção de QTL, seguiu-se o mesmo processo para simulação dos genomas e dos compostos, porém com algumas particularidades. Foram simulados genomas com dez grupos de ligação, devendo ser ressaltado que em cada genoma somente o grupo de ligação 1 foi simulado com cinco locos marcadores. Nestes grupos também foram distribuídos 200 genes controladores de características quantitativas. Porém, somente no grupo de ligação 1 foram alocados QTL e 19 genes de efeito menor. Em cada um dos demais grupos de ligação foram alocados mais 20 genes de efeito menor. Na análise de QTL, incluindo detecção e estimação dos efeitos de substituição e desvios de dominância, foram idealizadas três características quantitativas, com um QTL controlando cada característica. As características foram: produção (g/espiga), com direção de dominância positiva; capacidade de expansão (ml/g), com direção de dominância bidirecional; e crescimento líquido relativo (%), com direção de dominância negativa. Neste trabalho, foi empregado o método das marcas simples com base em análise de regressão. A avaliação do poder de detecção de QTL e os efeitos de substituição foram obtidos por regressão linear, enquanto os efeitos de dominância foram obtidos por regressão polinomial, incluindo o termo quadrático. No mapeamento de marcas moleculares, verificou-se que os grupos de ligação foram recuperados de forma satisfatória, com algumas exceções. Maior viés na recuperação dos genomas paramétricos foi verificado à medida que se diminuiu a diferença de frequência de mesma marca entre os genitores de 1,0 para 0,7. Com base no percentual de marcas ordenadas corretamente o mapeamento foi eficiente, apresentando valores superiores a 86%. Diante dos resultados obtidos concluiu-se que a população mais eficiente para o mapeamento é a equivalente a uma F_2 . Quanto à detecção de QTL, de modo geral, o poder de detecção foi satisfatório, independentemente do caráter idealizado. A análise do efeito de substituição de marca revelou maior magnitude de efeito para as marcas mais próximas ao QTL. Em relação ao teste de dominância nos locos marcadores, foi verificada eficiência na detecção de dominância apenas nos compostos obtidos do cruzamento entre genitores com diferença de frequência de mesma marca maior ou igual a 0,9 e com tamanho de amostra igual 1.000. Avaliando o sinal do efeito de dominância, para a característica produção, verificou-se que os

efeitos estimados nos locos marcadores foram consistentes em revelar a direção de dominância no loco do QTL, com algumas exceções quando dominância completa. Para característica capacidade de expansão, em virtude de ter sido considerada sob direção de dominância bidirecional, não foi possível a determinação de quais marcas têm efeito positivo ou negativo. Para a característica crescimento líquido relativo, as estimativas dos efeitos de dominância nos marcadores revelaram, quase sem exceção, a direção de dominância no loco do QTL. É importante ressaltar que os resultados obtidos neste trabalho, para o mapeamento de QTL, são válidos para QTLs que expliquem entre 10 e 30% a variação fenotípica do caráter.

ABSTRACT

SILVA, Admilson da Costa e, M.Sc., Universidade Federal de Viçosa, March 2008. **Constructing genetic map and indentifying QTL based on linkage disequilibrium**. Advisor: José Marcelo Soriano Viana. Co-advisors: Cosme Damião Cruz and Fabyano Fonseca e Silva.

In the study of genetic linkage between two marker loci or between markers and a QTL it is usually necessary to have appropriate designed populations (experimental populations). However, for some species such designed populations could be hard to obtain. In this scenario, the linkage disequilibrium (LD) has been used to build genetic map and to identify QTL. In this work, we used simulation to evaluate the efficiency of LD to construct genetic maps (LD-genetic map) of molecular markers as well as to map QTL (LD-QTL mapping) in non-structured families of outbreed populations. As for the LD-genetic map study, we designed genomes with high saturation level (markers apart from each other less than 5 cM), and random saturation level (markers apart from each other ranging from 0 to 20 cM), five linkage groups and 20 markers per linkage group. We simulated compounds of second generation with 200 and 1000 individuals, derived from the mating of two parental populations of size 200, under four levels (1.0, 0.9, 0.8 and 0.7) of difference in the frequency of each marker (marker-frequency difference) between parental populations. These compounds were used to obtain the

sampled genomes, which were compared with the designed genomes to measure the efficiency of LD to construct genetic map. In the LD-QTL mapping study, we adopted, with slightly modifications, the steps to obtain the designed genome and the compounds as described above. We designed genomes with ten linkage groups, each with 20 markers, but for linkage group 1 which had just five markers. We distributed 200 genes controlling quantitative traits along the ten linkage groups. However, only the linkage group 1 had a major QTL along with 19 minor genes. All other linkage groups had 20 minor genes each. In the LD-QTL mapping study, for the detection, estimation of substitution effects and of dominance, we idealized three quantitative traits, each being affected by a major QTL. The trait yield (g/ear) had a QTL with positive dominance; the trait expansion capacity (mL/g) had a QTL with bidirectional dominance; and the trait relative liquid growth (%) had a QTL with negative dominance. We used the method of single marker analysis based on regression to model the QTL effects. The analyses of power for QTL identification and substitution effect estimation were done by linear regression. On the other hand, the analysis of dominance effect was done by polynomial regression with the quadratic term included in the model. In the LD-genetic map study the linkage groups were satisfactorily recovered, with only few exceptions. The bias in recovering the designed genome increased with the decreasing of the marker-frequency difference between parental populations from 1.0 to 0.7. Based on the percentage of correct ordered markers in the genome, we found that the LD-genetic map was efficient, with correct ordering of markers superior to 86%. Our results support that the most efficient population for LD-genetic map was the one equivalent to an F_2 intercross population. With respect to the QTL detection, in general, the power of detection was satisfactory, independent of the quantitative trait simulated. The results of effect substitution of marker revealed that markers closer to the QTL had bigger effects. The detections of dominance effects were efficient only on the compounds from the mating of parental populations with marker-frequency difference superior than 0.9 and sample size of 1.000. The results of dominance direction analyses on the markers for the trait yield were efficient in determining the direction of dominance of QTL, with few exceptions where complete dominance was present. For the trait expansion capacity, the specification of which markers had positive or negative effect was not possible

because of the bidirectional nature of the simulated dominance effects. For the trait liquid growth the estimated effects of dominance at markers often revealed the correct dominance effects direction of the QTL. It is important to point out that the results of our LD-QTL mapping study were based on traits in which the phenotypic variance explained by the QTL is between 10 and 30% of the total phenotypic variance.

1- INTRODUÇÃO

O desenvolvimento e o uso de marcadores moleculares para detecção e exploração de polimorfismos de DNA em plantas e animais são uns dos mais significantes avanços na área de biologia molecular e biotecnologia. Os dois fenômenos envolvidos com a geração de polimorfismos genotípico, detectados por marcadores moleculares, são a mutação e a recombinação. A detecção de ligação e o histórico do polimorfismo de DNA têm sido o centro de diversos estudos com marcadores moleculares (NORDBORG e TAVARÉ, 2002). Entretanto, para o estudo de ligação é preciso realizar cruzamentos apropriados antes do desenvolvimento do mapa das populações. Esta é, em alguns casos, uma séria limitação ao uso de marcadores moleculares, pois os cruzamentos podem ser de difícil realização, como nas espécies perenes, e, ou, a população de mapeamento a ser utilizada pode ser muito pequena, com apenas dois alelos por loco amostrados.

Assim, métodos alternativos têm sido desenvolvidos e usados para estudar o fenômeno da ligação e recombinação e o histórico de mutações da população. O método com base na quantificação do desequilíbrio de ligação (LD) por análise de associação tem recebido, nos últimos anos, atenção especial dos geneticistas. Esta metodologia tem o potencial não apenas de

identificar e mapear QTL (*Quantitative Trait Loci*), mas também de identificar as causas de polimorfismos de dentro do gene, que seja responsável pela diferença entre dois fenótipos (PALAISA et al., 2003).

A construção de mapas genéticos é extremamente importante no melhoramento genético de plantas e animais. Estes mapas possibilitam a cobertura completa e a análise de genomas, a decomposição de caracteres complexos em seus componentes mendelianos simples e a localização de regiões do genoma responsáveis pelo controle da expressão de caracteres importantes, sejam eles qualitativos ou quantitativos (SILVA, 2005).

No melhoramento de plantas o mapeamento por LD tem várias vantagens sobre a análise de ligação clássica, usando populações segregantes. Em primeiro lugar, a ampla variabilidade genética na população, de origem mais representativa, pode ser incluída nas análises. Segundo, o mapeamento pode atingir uma resolução mais alta. Terceiro, podem ser utilizados dados fenotípicos de vários experimentos armazenados em bancos de dados, juntamente com dados moleculares (KRAAKMAN et al., 2004).

Populações derivadas de cruzamentos controlados ou endogâmicas como F_2 , retrocruzamentos, duplo-haplóides e linhas endogâmicas recombinantes (RILs – *Recombinant Inbred Lines*) têm sido as principais populações utilizadas na obtenção de mapas de ligação e no mapeamento de QTL. Populações exogâmicas, estruturadas em famílias, também têm recebido atenção, devendo ser ressaltado que cada uma tem suas vantagens e desvantagens.

Em populações não-naturais, onde não existe a estruturação em famílias, não é possível a realização de trabalhos semelhantes aos que geralmente são feitos em populações oriundas de cruzamentos controlados, que se baseiam apenas na análise de ligação clássica, ou seja, na frequência de recombinação gênica. Portanto, estudos com base em LD são de extrema importância na obtenção de mapas e detecção de QTL nessas populações.

Diante do exposto, este trabalho, realizado por meio de simulação de dados, teve como objetivos avaliar a eficiência do mapeamento de marcas moleculares e identificar o QTL em populações não-endogâmicas e não-estruturadas em famílias, com base no desequilíbrio de fase gamética.

2- REVISÃO DE LITERATURA

2.1- Desequilíbrio de Ligação

O termo desequilíbrio de ligação (*linkage disequilibrium-LD*), também conhecido como desequilíbrio de fase gamética (FLINT-GARCIA et al., 2003), é normalmente confundido com mapeamento de associação (*association mapping*). Entretanto, enquanto mapeamento de associação refere-se à associação significativa do marcador molecular com a característica fenotípica, LD refere-se à associação não-aleatória entre dois marcadores ou genes / QTL. Assim, o mapeamento de associação é atualmente um dos vários usos do LD. No sentido estatístico, associação se refere à covariância entre o marcador polimórfico e a característica de interesse, enquanto o LD representa a covariância entre o polimorfismo exibido entre dois marcadores moleculares/genes (GUPTA et al., 2005). Os termos ligação e LD também são freqüentemente confundidos. Embora sejam relacionados, eles são termos distintos. Ligação refere-se à herança correlacionada dos locos através de uma conexão física em um cromossomo, enquanto LD refere-se à correlação entre alelos na população. De forma simplificada, LD é a associação não-aleatória de alelos de locos diferentes. A confusão ocorre porque uma forte ligação pode resultar em altos níveis de LD (FLINT-GARCIA et al., 2003).

Diferentes metodologias para estimar os níveis de LD em plantas têm sido descritas em recentes revisões sobre LD em plantas (FLINT-GARCIA et al., 2003; GUPTA et al., 2005). Em algumas revisões são descritos os métodos disponíveis, a estatística utilizada para se testar a significância das medidas obtidas e as estimativas obtidas que envolvem locos multialélicos e condições multilocos (GORELICK e LAUBICHLER, 2004; GUPTA et al., 2005).

As duas estatísticas mais comuns para medir o LD são: r^2 e D' . Estas estimativas refletem diferentes aspectos do LD e comportam-se diferentemente sob condições variadas (FLINT-GARCIA et al., 2003). Dois locos, com dois alelos cada, ou seja, A e a, e B e b, com freqüências alélicas π_A , π_a , π_B e π_b , respectivamente, resultarão nas freqüências gaméticas π_{AB} , π_{Ab} , π_{aB} e π_{ab} . O componente básico para o cálculo de desequilíbrio é a diferença entre as freqüências esperada e observada dos gametas, dada por:

$$D_{ab} = (\pi_{AB} - \pi_A \cdot \pi_B)$$

A primeira medida do desequilíbrio (r^2), também descrita como Δ^2 , é dada por:

$$r^2 = \frac{(D_{ab})^2}{\pi_A \pi_a \pi_B \pi_b}$$

É conveniente considerar r^2 como o quadrado do coeficiente de correlação entre dois locos. Entretanto, a menos que os dois locos tenham freqüências alélicas idênticas, o valor 1 não é possível (FLINT-GARCIA et al., 2003).

Alternativamente, a estatística D' é calculada como:

$$|D'| = \frac{(D_{ab})^2}{\min(\pi_A \pi_b, \pi_a \pi_B)} \text{ para } D_{ab} < 0$$

$$|D'| = \frac{(D_{ab})^2}{\min(\pi_A \pi_B, \pi_a \pi_b)} \text{ para } D_{ab} > 0$$

O valor de D' é baseado nas freqüências alélicas observadas, e irá variar entre 0 e 1 se as freqüências alélicas diferirem entre os locos. D' poderá ser menor do que 1 apenas se todos os quatro possíveis gametas forem observados, conseqüentemente assumindo que eventos de recombinação ocorreram entre os locos (FLINT-GARCIA et al., 2003).

Vários fatores influenciam o LD. Alguns são responsáveis pelo aumento no LD, incluindo autofecundações, pequenos tamanhos de populações,

isolamento genético entre linhagens, subdivisão populacional, baixa taxa de recombinação, mistura populacional, seleção artificial e natural, dentre outros. Alguns outros fatores são responsáveis pela queda ou quebra do LD, incluindo acasalamento ao acaso, elevadas taxas de recombinação, elevadas taxas de mutações, dentre outros. Existem fatores que podem aumentar ou quebrar o LD, ou podem aumentar o LD entre determinado par de alelos e diminuir o LD entre outros pares de alelos. Por exemplo, a mutação pode romper o LD entre pares de alelos que envolvem alelos selvagens, e promover LD entre os pares de alelos dos mutantes envolvidos. Outros fatores que afetam o LD, incluindo estrutura populacional, epistasia e conversão gênica, não têm recebido atenção desejada nas revisões realizadas (GUPTA et al., 2005).

De forma resumida, os fatores explicados por Flint-Garcia et al. (2003) que afetam o LD são apresentados a seguir.

A frequência alélica e a recombinação entre locos, assim como as mutações, afetam o LD. A recombinação é o principal fenômeno que diminui o LD intracromossômico, enquanto o LD intercromossômico diminui com a distribuição independente. O tamanho populacional também tem importante papel, pois em pequenas populações os efeitos da deriva genética resultam em perda consistente de combinações alélicas raras, o que aumenta os níveis de LD.

Os sistemas de acasalamento e misturas também podem fortemente influenciar o LD. Geralmente o LD diminui mais rapidamente em espécies alógamas, comparadas com espécies autógamias. Isto porque a recombinação é menos efetiva em espécies que se autofecundam, em que os indivíduos são mais semelhantes por serem homozigotos, do que em espécies de fecundação cruzada. A mistura é fluxo de genes entre indivíduos geneticamente distintos, seguida por intercruzamentos. Ela resulta na introdução de genes de diferentes ancestrais e mudança de frequências alélicas. Porém, em geral, o LD resultante cai rapidamente com os cruzamentos aleatórios.

O LD pode também ser criado em populações que tiveram recentemente seu tamanho populacional reduzido devido ao afunilamento genético (*bottleneck*). Durante o afunilamento, apenas poucos alelos combinados são passados para gerações futuras, o que pode gerar um LD considerável.

A seleção, que causa afunilamento genético para locos específicos, também pode criar LD entre o alelo selecionado e o loco ligado. Além disso, a seleção a favor ou contra um fenótipo determinado por dois genes não-ligados pode criar um LD, embora os locos não estejam fisicamente ligados. Finalmente, o fluxo gênico entre populações pode introduzir novas combinações cromossômicas e diferentes frequências alélicas, resultando em LD.

2.1.1- Desequilíbrio de ligação em genômica de plantas

Segundo Flint-Garcia et al. (2003), pesquisas sobre desequilíbrio de ligação têm sido realizadas mais extensivamente em sistemas animais e de forma mais restrita em plantas. A maioria das pesquisas sobre LD em plantas foi conduzida em milho e *Arabidopsis*.

O desequilíbrio de ligação pode ser utilizado para vários propósitos em pesquisas genômicas em plantas. Uma das maiores utilizações atuais e futuras do LD em plantas provavelmente será no estudo da associação marcador - característica (sem o uso de mapeamento de populações), seguido de seleção assistida por marcadores. Outras utilizações importantes encontram-se no estudo de diversidade genética em populações naturais e coleções de germoplasma, em estudos de genética de populações e em programas de melhoramento de plantas (GUPTA et al., 2005).

A associação entre característica-marcador em plantas é geralmente feita por análise de ligação, utilizando análise de regressão simples e mapeamento de QTL por intervalo simples ou composto. Algumas limitações desses métodos têm sido superadas com o mapeamento de associação com base no LD. Para o estudo de associação característica – marcador usando o LD, as metodologias diferem para caracteres quantitativos e caracteres discretos (qualitativos), embora ocasionalmente os caracteres quantitativos possam ser tratados como caracteres qualitativos. Dois métodos têm sido comumente usados para caracteres discretos em humanos, para mapeamento de genes controladores de doenças: (i) case-control (CC) e (ii) teste de transmissão de desequilíbrio (TDT) (SPIELMAN et al., 1993). Metodologias similares, mas não idênticas, têm sido usadas em plantas. Como exemplos de

estudos que envolvem características qualitativas em plantas, podem-se citar dois trabalhos conduzidos em milho. Um foi realizado por Palaisa et al. (2003), buscando associação entre polimorfismo do gene *y1* e coloração do endosperma. A metodologia usada nesse estudo é comparável à usada nos estudos de CC em humanos. Outro foi conduzido por Kumar et al. (2004), no qual 200 famílias de irmãos completos foram usadas para estudar associação marcador – característica em pinus. Nesse estudo os genótipos dos parentais foram também considerados durante as análises, e esse método pode ser comparado com TDT em humanos.

O uso do LD para mapeamento de QTL para caracteres quantitativos é mais problemático, porém é também mais recompensador, uma vez que ele permite localizar a posição do QTL que controla a característica de interesse de forma mais precisa. Mackay (2001a) e Glazier et al. (2002) sugerem que se utilize análise de ligação para uma localização preliminar do QTL e depois se use o LD para localização mais precisa.

2.2- Marcadores moleculares

O termo marcador genético corresponde a uma característica do organismo que pode ser facilmente detectada a olho nu, ou com a ajuda de algum aparato tecnológico, e que co-segrega com genes de interesse. Uma característica, para ser útil como marcador, deve evidenciar diferenças entre os indivíduos analisados e, além disso, ser reproduzida com precisão na prole (SILVA, 2005). Esses marcadores podem ser classificados em quatro tipos: morfológicos, citológicos, bioquímicos e moleculares (BORÉM e MIRANDA, 2005).

Até meados da década de 1960, nos estudos de genética e melhoramento eram utilizados basicamente marcadores morfológicos, em geral fenótipos de fácil identificação visual, como nanismo, cor de pétala, morfologia foliar, entre outros. Estes marcadores contribuíram bastante para o desenvolvimento teórico da análise de ligação gênica e para a construção das primeiras versões de mapas genéticos. Contudo, devido ao reduzido número de marcadores morfológicos polimórficos, a probabilidade de encontrar associações entre esses marcadores e características de interesse era

reduzida, limitando sua utilização em análises genéticas (FERREIRA e GRATTAPAGLIA, 1998; GUIMARÃES e MOREIRA, 2005).

Com o advento das técnicas de biologia molecular tornou-se possível a manipulação do DNA, que culminou no surgimento dos vários tipos de marcadores moleculares disponíveis atualmente. Se um marcador apresentar comportamento de acordo com as leis básicas de herança, enunciadas por Mendel, ele será um “marcador genético” e poderá ser monitorado ao longo das gerações (FERREIRA e GRATTAPAGLIA, 1998).

Os marcadores moleculares apresentam algumas vantagens em relação aos marcadores morfológicos, como: o nível de polimorfismo é geralmente alto para cada loco estudado, facilitando o desenvolvimento de mapas genéticos a partir de populações segregantes de cruzamentos específicos; e a neutralidade em relação aos efeitos de ambiente, com pouco ou nenhum efeito de epistasia ou pleiotropismo (FERREIRA e GRATTAPAGLIA, 1998; LANZA et al., 2000).

Os marcadores moleculares mais utilizados em estudos genéticos de plantas e melhoramento são RFLP (*Restriction Fragment Length Polymorphism*), microssatélites ou SSR (*Simple Sequence Repeats*), RAPD (*Random Amplified Polymorphic DNA*) e AFLP (*Amplified Fragment Length Polymorphism*). Um tipo mais recente, também bastante utilizado, são os marcadores baseados em polimorfismo de base única (SNPs - *Single Nucleotide Polymorphism*) (SOUZA, 2001; CAIXETA et al., 2006).

Durante algum tempo, os marcadores RFLP foram bastante utilizados para detectar polimorfismos na molécula de DNA. Como possuem expressão co-dominante, possibilitam a identificação de genótipos homozigotos e heterozigotos. Esse tipo de marcador tem sido utilizado na construção de mapas genéticos de diversas espécies, como cana (GUIMARÃES et al., 1999), milho (CARDINAL et al., 2001), feijão (MIKLAS et al., 2001) e soja (NJITI et al., 2002).

Microssatélites ou SSR consistem de pequenas seqüências com um a quatro nucleotídeos de comprimento, repetidas em tandem no genoma. Devido à expressão co-dominante desses marcadores e ao seu multialelismo, eles podem ser aplicados a todos os tipos de populações segregantes empregadas no mapeamento genético e estudos de ligação (FERREIRA e GRATTAPAGLIA, 1998). Esses marcadores têm sido amplamente utilizados

em análise genética, uma vez que se baseia na detecção da variação existente em locos de seqüências repetitivas (di-nucleotídeos AG/TC ou tri-nucleotídeos ATT/TAA, por exemplo). As seqüências que flanqueiam esses locos são mais conservadas, permitindo que *primers* específicos a elas sejam sintetizados e utilizados para amplificar especificamente essa região, via PCR. Os marcadores microssatélites são altamente informativos, permitindo detectar polimorfismo molecular com grande eficácia (LITT e LUTY, 1989). Em trabalhos como os de Brondani et al. (2001), sobre mapeamento em arroz; Sibov et al. (2003a), sobre construção de mapa genético em milho; Salla et al. (2002), sobre variabilidade genética em acerola; Liu et al, 2003, sobre diversidade genética em milho; Radmann et al. (2006), sobre caracterização e diversidade genética em morangueiro; Sibov et al. (2003b), Garcia (2003), Mangolin et al. (2004), Jardim et al. (2005) e Lima (2006), sobre mapeamento de QTL em milho; Santos et al. (2006), sobre identificação de QTL em soja; e Mathias et al. (2007), sobre avaliação de germoplasma de batata, foram empregados marcadores microssatélites.

Os marcadores RAPD requerem uma pequena quantidade de DNA para análise, além do fato de um simples *primer* poder revelar muitos locos de uma só vez, cada loco correspondendo a uma região diferente do genoma analisado. Por serem dominantes, não permitem distinguir o genótipo homozigoto dominante do heterozigoto. Esses marcadores também têm sido utilizados por diversos pesquisadores no mapeamento genético de feijão (FALEIRO et al., 2003) e de fruteiras (OLIVEIRA et al., 2004); na identificação de linhagens de feijão resistentes a doenças (ALZATE-MARIN et al., 2003); e na determinação da diversidade genética em gergelim (ARRIEL et al., 2006) e laranja (BASTIANEL et al., 2006).

O AFLP é outro tipo de marcador amplamente empregado em análises genéticas. Estes marcadores apresentam como vantagem o grande poder de detecção de variabilidade genética, uma vez que a técnica explora polimorfismos de restrição e de amplificação. Dessa forma, é resolvido um grande número de fragmentos polimórficos em um único gel. Como os *primers* utilizados nas etapas de amplificação são longos, em torno de 20 pb, a especificidade das reações é significativamente aumentada, quando comparada com o RAPD. Assim, o AFLP alia a vantagem de explorar regiões

genômicas arbitrárias, sem a necessidade do conhecimento prévio das seqüências do DNA, com a elevada especificidade da técnica de PCR (LANZA et al., 2000). Outro ponto positivo é que são distribuídos aleatoriamente pelo genoma, fazendo com que tenham grande potencial para assistir diferentes áreas da genética e do melhoramento de plantas, como monitoramento da herança de características de interesse agrônômico, diagnóstico de doenças, análise de pedigree e de diversidade genética, assim como na seleção assistidas por marcadores (CAIXETA et al., 2006).

Os marcadores AFLP e RAPD podem ser clonados e convertidos em marcadores específicos, denominados STS e SCAR respectivamente. Esses marcadores, geralmente, são desenvolvidos a partir de seqüências de genes ou de marcadores associados a características de interesse, como os marcadores SCARF10 e SCARFBA8, ligados a genes de resistência à ferrugem do feijoeiro (FALEIRO et al., 2003).

Os SNPs ou polimorfismos simples de nucleotídeos são o mais recente tipo de marcador molecular conhecido. Sua existência foi evidenciada no decorrer dos projetos de seqüenciamento de genomas e de genes expressos, ao se observar um grande número de variações de nucleotídeos. Sua determinação corresponde a posições onde existe uma alternância dos nucleotídeos A, C, G, e T, em uma freqüência alélica mínima de 1%, em uma dada população. Esses polimorfismos são analisados em grandes populações, visando encontrar algum tipo de associação entre um polimorfismo e um determinado fenótipo de interesse. Na área vegetal, o seqüenciamento em larga escala tem permitido evidenciar a presença de SNPs em espécies cultivadas (CAIXETA et al., 2006). Esses marcadores têm sido como uma fonte inesgotável de polimorfismos para uso no mapeamento genético de alta resolução e para estudos em genética de associação (RAFALSKI, 2002). Trabalhos como os de Ching et al. (2002) e Palaisa et al. (2003), na estimação da taxa de LD em milho, e ZHU et al., (2003), em soja, foram utilizados marcadores SNPs.

Esses vários tipos de marcadores moleculares se diferenciam entre si quanto à sua habilidade em detectar polimorfismo, ao custo de aplicação, facilidade de uso e à consistência de resultados (BORÉM, 2006). A utilização de uma ou outra técnica de marcadores moleculares deve ser guiada pelo

problema que efetivamente se apresenta no melhoramento de uma cultura ou análise genética, levando em conta as vantagens e limitações de cada classe de marcador disponível. Não se deve dirigir somente pelo fato de esta ou daquela tecnologia ser mais sofisticada ou estar em voga no momento (FERREIRA e GRATTAPAGLIA, 1998).

2.3- Mapas genéticos

Mapas de ligação ou genéticos são representações gráficas dos cromossomos de uma espécie, obtidos por meio de análises estatísticas. Um mapa de ligação é composto por vários grupos de ligação, sendo o número deles dependente da espécie em estudo. Dessa forma, um grupo de ligação contém locos ordenados, podendo estes ser genes (marcadores morfológicos ou isoenzimáticos) ou não (marcadores moleculares), alinhados ao longo de um mesmo cromossomo, indicando as distâncias relativas entre eles. Mapas genéticos são importantes em estudos genéticos, pois estruturam a base do conhecimento genômico, a partir do qual podem ser esclarecidos vários aspectos referentes ao mapeamento de caracteres de interesse, à manipulação genética, à evolução das espécies e à expressão e função gênica (LYNCH e WALSH, 1998).

O conceito utilizado na construção de mapas de ligação baseia-se na hipótese de que a co-transmissão de dois genes reflete a proximidade entre eles, uma vez que a probabilidade de ocorrer *crossing-over*, ou seja, de ocorrer recombinação, é tanto menor quanto mais próximos estiverem um do outro. As freqüências de recombinações na amostra são medidas, e determina-se o grau de ligação entre os genes em estudo. A freqüência de recombinação entre dois locos é função da distância entre eles. Quanto maior a distância, maior a probabilidade de ocorrência de recombinação entre os locos. Assim, a freqüência de recombinação pode ser utilizada como uma medida de distância entre os genes (LIU, 1998).

As freqüências de recombinação são estimativas das probabilidades de ocorrência de *crossing-over* e, como tais, não possuem a propriedade de serem aditivas. Outro aspecto, o fenômeno de interferência, isto é, a ocorrência de um *crossing-over* afetando a ocorrência de um segundo em regiões adjacentes, torna menos precisa a utilização da freqüência de recombinação

para estimar a distância entre os genes. Diante disso, faz-se necessária a transformação de escala para o cálculo das distâncias entre genes, para tornar as frequências de recombinação em medidas aditivas de distância. Um significativo número de funções de mapeamento, que relacionam medidas de distâncias às frequências de recombinação, foi desenvolvido. Contudo, as mais utilizadas no mapeamento genético são as funções de Haldane e Kosambi. A primeira supõe ausência de interferência e a segunda considera uma interferência moderada (LANZA et al., 2000; SCHUSTER e CRUZ, 2004).

Para frequências de recombinação (inferior a 20%), ambas as funções de mapeamento apresentam valores similares aos da frequência de recombinação. No entanto, à medida que a distância entre dois locos aumenta, ou seja, com o aumento da frequência de recombinação, mais discrepantes tornam-se os valores obtidos pelas funções de Kosambi e Haldane, entre si e em relação à frequência de recombinação (SCHUSTER e CRUZ, 2004).

2.4- Populações de mapeamento de QTL

Para o mapeamento genético, diferentes tipos de populações são empregadas. Estas populações são comumente chamadas de populações de mapeamento. Elas se dividem em populações endogâmicas ou populações derivadas de cruzamentos controlados e em populações exogâmicas ou populações naturais.

As populações endogâmicas F_2 , de retrocruzamento e linhagens recombinantes (*Recombinant Inbred Lines*, RILs), são os tipos mais utilizados para construção de mapas moleculares. Elas são derivadas do cruzamento de linhas puras, que originam uma geração F_1 , que é autofecundada para obter a geração F_2 ou retrocruzada com um dos parentais para produção de uma geração de retrocruzamento (FERREIRA e GRATTAPAGLIA, 1998; BRONDANI et al., 2003). As populações de RILs (BURR et al., 1988) são obtidas de autofecundações sucessivas (oito a nove ciclos) de uma F_2 derivada de progenitores endogâmicos, obtendo-se, ao final do processo, uma série de linhagens de cada planta F_2 original. Também podem ser utilizadas populações F_n ($n= 3, 4, \dots, \infty$) e duplo-haplóides (DH) (BRONDANI et al., 2003; PEREIRA e PEREIRA, 2006). A escolha da população deve levar em conta os objetivos do pesquisador, além do tempo e dos recursos disponíveis (LANZA et al., 2000;

SHUSTER e CRUZ, 2004). É importante ressaltar que nos indivíduos F_1 o desequilíbrio de ligação é máximo, e os estudos das populações derivadas a partir desses indivíduos procuram explorar este desequilíbrio (SCHUSTER e CRUZ, 2004).

O desequilíbrio pode ser decorrente de diversas causas, incluindo seleção e deriva genética. Entretanto, a causa predominante em populações segregantes, derivadas de cruzamentos entre linhagens, é a ligação física dos locos. Este tipo de desequilíbrio de ligação é bastante elevado nas populações de cruzamentos controlados. Conseqüentemente, a capacidade de se detectar a ligação entre os dois locos fisicamente ligados também é alta (LYNCH e WALSH, 1998; COELHO, 2000). Na análise de QTL o desequilíbrio de ligação entre o marcador genético e o QTL é a condição essencial para sua detecção e localização.

Nas espécies florestais e outras perenes ou alógamas, a de obtenção e manipulação de linhagens endogâmicas a partir de cruzamentos controlados não é simples. Nestes casos deve-se, então, voltar para a análise e utilização de populações exogâmicas (LYNCH e WALSH, 1998). Estas populações podem estar estruturadas em famílias de meios-irmãos ou de irmãos completos, ou sem estrutura de famílias.

As populações exogâmicas são também de interesse, por razões outras que as limitações experimentais. Por exemplo, os QTLs detectados por cruzamentos controlados, geralmente, representam diferenças fixas entre linhas, ou muitas vezes entre espécies, e a relevância destes resultados para QTLs segregando dentro de populações permanece incerta. Essa fundamental distinção de que os cruzamentos controlados detectam QTLs responsáveis por diferenças entre populações, enquanto os cruzamentos exogâmicos detectam QTL responsáveis por variação dentro de populações, torna essas abordagens mais complementares do que competitivas (LYNCH e WALSH, 1998).

A utilização da variação dentro da população, em oposição às diferenças fixas entre populações, resulta em redução significativa do poder de detecção de QTL. Com linhagens endogâmicas, todos os parentais F_1 têm genótipos idênticos (incluindo a mesma fase de ligação), portanto todos os indivíduos são informativos e o desequilíbrio de ligação é maximizado. Adicionalmente, os efeitos dos QTLs são expressos como médias (o valor médio de cada genótipo

do QTL). Em contrapartida, os efeitos dos QTLs são expressos como variâncias em cruzamentos exogâmicos. Uma vez que a variância é estimada com precisão menor do que a média, as estimativas para populações exogâmicas são tidas como menos precisas (LYNCH e WALSH, 1998).

2.5- Detecção de QTL

Mapear QTL consiste em analisar regiões cromossômicas que contenham genes (ou locos) que influenciam a variação de caracteres quantitativos. Análises desse tipo permitem identificar tanto o número como a distribuição dos QTL no genoma, além de identificar o tipo e a magnitude de seus efeitos (LIU, 1998; LYNCH e WALSH, 1998).

Em geral, no mapeamento de QTL a herança complexa dos caracteres quantitativos é relacionada à herança simples de caracteres qualitativos conhecidos, com base em marcadores genéticos. Diversos métodos estatísticos permitem quantificar essa relação, analisando a associação entre a variação de dados fenotípicos contínuos e o padrão de segregação de marcadores em populações segregantes. Associações significativas entre os fenótipos e os marcadores indicam a existência de QTLs localizados próximos a esses locos marcadores e permitem que sejam feitas inferências sobre os QTL (LIU, 1998). A eficiência do mapeamento de QTL depende do método de mapeamento empregado, do tipo de população de mapeamento e da saturação do mapa de ligação, além do tamanho da população e da herdabilidade do caráter a ser investigado (MACHAY, 2001b).

A seguir são apresentados alguns trabalhos sobre mapeamento de QTL em milho. Nesses trabalhos constou-se variação na quantidade de caracteres para os quais foram mapeados QTLs, tipo e número de marcadores empregados, assim como métodos de análise empregados.

Austin e Lee (1998) estudaram a consistência de QTL para produção de grãos e quatro caracteres significativamente correlacionados à produção (peso de sementes, número de espigas por planta, comprimento e diâmetro da espiga), em 185 famílias $F_{6:7}$ obtidas da população Mo17 x H99, avaliadas em 1993 e 1994, em Iowa, EUA. Também compararam os QTLs dessas famílias com famílias $F_{2:3}$ obtidas da mesma população e no mesmo local, em 1989 e 1990. Para isso, empregaram o método de mapeamento por intervalo

composto em um mapa genético constituído de 87 marcadores do tipo RFLP e um marcador morfológico. Nas famílias $F_{6:7}$, encontraram entre 10 e 20 QTLs para os caracteres correlacionados e oito QTLs para produção de grãos, considerando os dois ambientes. Os oito QTLs foram responsáveis por 27% da variância fenotípica e 68% da variância genotípica do caráter. Destes, apenas um foi considerado comum aos dois ambientes e apenas três apresentaram efeito aditivo. A comparação entre os dois grupos de famílias resultou em três QTLs comuns, localizados nos cromossomos 3, 6 e 8.

Sibov et al. (2003a) mapearam QTLs associados à produção de grãos, à altura de planta, à altura de espiga e ao teor de umidade nas progênies $F_{2:3}$ derivadas de plantas F_2 autofecundadas, utilizando o mesmo método de mapeamento do trabalho citado anteriormente. Foram mapeados quatro QTLs para produção de grãos, quatro para altura de planta e cinco para altura de espiga. Os QTLs mapeados encontram-se distribuídos pelos cromossomos 1, 2, 7, 8 e 9. Os QTLs, em conjunto, foram responsáveis por aproximadamente 32,73, 24,76 e 20,91% da variação fenotípica para os caracteres produção de grãos, altura de planta e altura de espiga, respectivamente. Os tipos de ações gênicas mais freqüentes encontradas foram dominância parcial e sobredominância. Alguns QTLs para altura da planta e de espiga foram mapeados próximos aos QTL para produção de grãos, indicando ligação ou efeitos pleiotrópicos nessas regiões.

Mangolin et al. (2004) mapearam QTLs associados ao teor de óleo em sementes $F_{2:3}$. Para isso, utilizaram 75 marcadores SSR e empregaram o método de mapeamento por intervalo composto. Foram mapeados 13 QTLs, distribuídos em oito cromossomos, responsáveis por 26,64% da variância genotípica. A contribuição principal para o caráter foi proveniente dos QTLs mapeados nos cromossomos 1, 5 e 6. Verificou-se o predomínio de QTLs cujas ações gênicas foram de dominância parcial, mas ações aditiva e de sobredominância também foram observadas. Oito dos 13 QTLs mapeados já foram reportados em populações de milho temperado.

Lima (2006) mapeou QTLs associados com produção de grãos, número de espigas por planta, acamamento, altura de planta, altura de espiga e número de folhas, em 256 progênies $F_{2:3}$ derivadas do cruzamento de duas linhagens de milho tropical, avaliadas em cinco ambientes. Na construção do

mapa genético foram utilizados 139 marcadores microssatélites. O mapeamento QTL e a detecção da interação QTL x ambiente foram realizados com o emprego da análise de mapeamento em ambientes múltiplos (mCIM). Foram mapeados 16 QTL para produção de grãos, oito para espiga por planta, seis para acamamento, seis para altura de planta, seis para altura de espiga e dois para número de folhas, distribuídos em dez grupos de ligação. Os efeitos genéticos dos QTLs variam em sinal e magnitude, indicando que cada QTL contribui de forma particular para a expressão dos caracteres. A maioria dos QTL mapeados apresentou ação gênica sobredominante e interação QTL X ambiente significativa.

2.6- Métodos de mapeamento

De modo semelhante às metodologias de mapeamento genético, o princípio básico que fundamenta as análises de mapeamento de QTL é a existência de desequilíbrio de ligação decorrente da redução da frequência de recombinação entre genes situados próximos entre si (COELHO, 2000). A maximização do desequilíbrio de ligação, aliada à possibilidade de construir mapas de ligação saturados, permite encontrar associações significativas entre locos marcadores e QTL (LIU, 1998; LYNCH e WALSH, 1998).

As metodologias estatísticas empregadas no mapeamento de QTL definem como utilizar o desequilíbrio de ligação e considerar a ligação entre os locos marcadores e os QTLs. Dentre os diversos métodos de mapeamento propostos, destacam-se: a análise de marcas simples (STUBER et al., 1987), o mapeamento por intervalo (LANDER e BOTSTEIN, 1989), o mapeamento por intervalo composto (JANSEN e STAM, 1994; ZENG, 1994) e o mapeamento de múltiplos intervalos (KAO et al., 1999). O método de análise de marcas simples normalmente utiliza algum modelo linear, enquanto os métodos de mapeamento por intervalo, por intervalo composto e de múltiplos intervalos utilizam métodos de máxima verossimilhança, baseados na informação da distribuição conjunta dos marcadores e dos QTLs (LIU, 1998; LYNCH e WALSH, 1998).

O método de análise de marca simples é utilizado para uma abordagem inicial no mapeamento de QTL. Nessa análise, a distribuição dos valores do caráter é examinada separadamente, para cada loco marcador. Assim, cada

teste de associação caráter-marcador é realizado independentemente da informação dos outros locos marcadores. Para um cromossomo com n marcas, n testes são realizados. A estratégia para detectar QTL, nesta análise, é fundamentada na comparação entre os valores médios da característica associados aos diferentes genótipos marcadores. Uma diferença significativa entre as médias das classes genotípicas dos marcadores indica a possibilidade da presença de QTL associado ao marcador. Quando o delineamento experimental apresenta apenas duas classes genotípicas, como em retrocruzamentos, a comparação entre eles pode ser realizada através do teste t . Havendo mais de duas classes genotípicas, as comparações são realizadas utilizando testes padrões da teoria de modelos lineares, como teste F em análises de variância ou modelos de regressão linear (LIU, 1998; LYNCH e WALSH, 1998).

3- MATERIAL E MÉTODOS

3.1- Aspectos teóricos

3.1.1- Análise de desequilíbrio de fase gamética

Embora, em geral, não reconhecido como o fundamento do mapeamento de marcas moleculares e da identificação e localização de *Quantitative Trait Loci*, o desequilíbrio de fase gamética é o fenômeno que permite a elaboração de mapas de marcas moleculares e a análise de QTL. No caso do delineamento de F_2 , a medida do desequilíbrio entre dois segmentos de DNA (genes e, ou, marcas) é

$$\Delta = \left(\frac{1 - 2r}{4} \right), \text{ em caso de aproximação, e}$$

$$\Delta = - \left(\frac{1 - 2r}{4} \right), \text{ em caso de repulsão,}$$

sendo r a frequência de gametas recombinantes em relação aos dois segmentos. Portanto, em relação a segmentos com distribuição independente ($r = 1/2$), ligados ou não, $\Delta = 0$.

No 'pool' gamético de uma população (geração -1), as probabilidades dos gametas AB, Ab, aB e ab são (KEMPTHORNE, 1957):

$$P_{11}^{(-1)} = p_a p_b + \Delta^{(-1)}$$

$$P_{10}^{(-1)} = p_a q_b - \Delta^{(-1)}$$

$$P_{01}^{(-1)} = q_a p_b - \Delta^{(-1)}$$

$$P_{00}^{(-1)} = q_a q_b + \Delta^{(-1)}$$

em que p e q são freqüências alélicas e $\Delta^{(-1)} = P_{11}^{(-1)}P_{00}^{(-1)} - P_{10}^{(-1)}P_{01}^{(-1)}$ é a medida do desequilíbrio de fase gamética.

As probabilidades dos genótipos na geração de acasalamento ao acaso (geração 0), a população de análise de desequilíbrio e a de análise de QTL são:

$$f_{22}^{(0)} = p_a^2 p_b^2 + 2p_a p_b \Delta^{(-1)} + \left[\Delta^{(-1)} \right]^2$$

$$f_{21}^{(0)} = 2p_a^2 p_b q_b + 2p_a (q_b - p_b) \Delta^{(-1)} - 2 \left[\Delta^{(-1)} \right]^2$$

$$f_{20}^{(0)} = p_a^2 q_b^2 - 2p_a q_b \Delta^{(-1)} + \left[\Delta^{(-1)} \right]^2$$

$$f_{12}^{(0)} = 2p_a q_a p_b^2 + 2(q_a - p_a) p_b \Delta^{(-1)} - 2 \left[\Delta^{(-1)} \right]^2$$

$$f_{11c}^{(0)} = 2p_a q_a p_b q_b + 2p_a p_b \Delta^{(-1)} + 2q_a q_b \Delta^{(-1)} + 2 \left[\Delta^{(-1)} \right]^2$$

$$f_{11r}^{(0)} = 2p_a q_a p_b q_b - 2p_a q_b \Delta^{(-1)} + 2q_a p_b \Delta^{(-1)} + 2 \left[\Delta^{(-1)} \right]^2$$

$$f_{10}^{(0)} = 2p_a q_a q_b^2 - 2(q_a - p_a) q_b \Delta^{(-1)} - 2 \left[\Delta^{(-1)} \right]^2$$

$$f_{02}^{(0)} = q_a^2 p_b^2 - 2q_a p_b \Delta^{(-1)} + \left[\Delta^{(-1)} \right]^2$$

$$f_{01}^{(0)} = 2q_a^2 p_b q_b - 2q_a (q_b - p_b) \Delta^{(-1)} - 2 \left[\Delta^{(-1)} \right]^2$$

$$f_{00}^{(0)} = q_a^2 q_b^2 + 2q_a q_b \Delta^{(-1)} + \left[\Delta^{(-1)} \right]^2$$

em que f_{ij} é a probabilidade do portador de i e j cópias dos segmentos de DNA A e B. Os índices c e r identificam os duplo-heterozigotos em fase de aproximação e repulsão, respectivamente (VIANA, 2004).

Então, assumindo que A e B são marcas moleculares bi-alélicas, de herança co-dominante, é possível estimar a medida do desequilíbrio entre elas e testar a hipótese de equilíbrio ($H_0: \Delta^{(-1)} = 0$). Analisando todos os possíveis pares de marcas pode-se formar um ou mais grupos de desequilíbrio. Em F_2 e em compostos derivados do cruzamento entre duas ou mais populações (de segunda geração), os grupos de desequilíbrio correspondem, teoricamente, aos grupos de ligação, o que viabiliza, embora com diferentes graus de eficiência, o mapeamento de marcas moleculares e a localização de um ou mais QTL.

Empregando o método da máxima verossimilhança, a função de verossimilhança é:

$$L(y | p_a, p_b, \Delta^{(-1)}) = \frac{N!}{N_{22}! \cdot N_{21}! \cdot \dots \cdot N_{00}!} \left[f_{22}^{(0)} \right]^{N_{22}} \left[f_{21}^{(0)} \right]^{N_{21}} \dots \left[f_{00}^{(0)} \right]^{N_{00}}$$

em que N_{ij} é o número de indivíduos portadores de i e j cópias dos segmentos de DNA A e B, e N é o número total de indivíduos genotipados.

Em razão da complexidade dessa função, os estimadores de máxima verossimilhança podem ser obtidos por *grid search* (procura pelo máximo), usando a função de verossimilhança ou a função suporte (logaritmo natural da função de verossimilhança). Para minimizar o tempo de processamento, pode-se primeiramente obter as estimativas de máxima verossimilhança das frequências das marcas moleculares, obtendo-se posteriormente, por *grid search*, a estimativa da medida do desequilíbrio.

Em relação a cada marca, tem-se (ELANDT-JOHNSON, 1971):

- estimador de máxima verossimilhança de p : $\hat{p} = (\text{frequência de homozigotos}) + (1/2) \cdot (\text{frequência de heterozigotos})$;
- limite inferior da variância de \hat{p} : $V(\hat{p}) \geq \frac{pq}{2N}$; e
- conteúdo médio de informação por observação: $\frac{I(\hat{p})}{N} \leq \frac{2}{pq}$.

Antes do teste de equilíbrio de fase gamética deve-se testar se as duas marcas estão em equilíbrio de *Hardy-Weinberg* (teste qui-quadrado, com um grau de liberdade). Para o teste de equilíbrio de fase gamética pode-se empregar teste qui-quadrado (com 4 graus de liberdade), LOD score ou teste da razão de verossimilhança, dados por (LIU, 1998):

$$\left(\chi_{\Delta=0}^2 \right) = \chi_1^2 - \chi_a^2 - \chi_b^2$$

sendo χ_1^2 o valor calculado ao admitir equilíbrio de *Hardy-Weinberg* para as duas marcas e equilíbrio de fase gamética, e χ_a^2 e χ_b^2 os valores calculados admitindo equilíbrio de *Hardy-Weinberg*.

$$\text{LOD} = \log_{10} \frac{L\left(y | \hat{p}_a, \hat{p}_b, \hat{\Delta}^{(-1)}\right)}{L\left(y | \hat{p}_a, \hat{p}_b\right)}$$

$$\text{LR} = 2 \cdot \ln \frac{L\left(y | \hat{p}_a, \hat{p}_b, \hat{\Delta}^{(-1)}\right)}{L\left(y | \hat{p}_a, \hat{p}_b\right)} \sim \chi^2 \quad c/1 \text{ g.l.}$$

Como apenas a estimação das distâncias entre as marcas possibilita a elaboração de mapas de marcas moleculares e a localização de QTL, a estimação de medidas de desequilíbrio e a formação de grupos de desequilíbrio não são tão relevantes, a não ser, como acontece em F_2 , que a distância seja uma função da medida do desequilíbrio e que os grupos de desequilíbrio sejam os grupos de ligação.

Como exemplos de populações alógamas em equilíbrio de *Hardy-Weinberg* e com desequilíbrio de fase gamética, podem ser citados compostos de duas ou mais populações (de segunda geração), sintéticos de duas ou mais linhagens (de segunda geração) e populações derivadas de cruzamentos entre híbridos simples, duplos ou triplos (de segunda geração). Considerando como população de mapeamento e de identificação de QTL a segunda geração de um composto de duas populações, o 'pool' gamético do híbrido interpopulacional (geração -1), em relação a duas marcas bi-alélicas, é:

$$P_{11}^{(-1)} = \left(\frac{p_a^1 + p_a^2}{2} \right) \left(\frac{p_b^1 + p_b^2}{2} \right) + \left(\frac{1-2r}{4} \right) (p_a^1 - p_a^2) (p_b^1 - p_b^2) = p_a p_b + \Delta^{(-1)}$$

$$P_{10}^{(-1)} = p_a \left(\frac{q_b^1 + q_b^2}{2} \right) - \Delta^{(-1)} = p_a q_b - \Delta^{(-1)}$$

$$P_{01}^{(-1)} = \left(\frac{q_a^1 + q_a^2}{2} \right) p_b - \Delta^{(-1)} = q_a p_b - \Delta^{(-1)}$$

$$P_{00}^{(-1)} = q_a q_b + \Delta^{(-1)}$$

sendo p^1 e p^2 as freqüências de mesma marca nas populações genitoras.

Então, a medida do desequilíbrio de fase gamética nessa população é função da freqüência de gametas recombinantes entre as marcas, e conseqüentemente da distância entre elas, e das diferenças de freqüências de mesma marca entre os genitores. O valor paramétrico revela que em um composto os grupos de desequilíbrio são os grupos de ligação, pois $\Delta = 0$ quando $r = 1/2$. Para todo $r < 1/2$, seu valor é máximo quando as diferenças de freqüências de mesma marca nas populações cruzadas são máximas (1 ou -1 para cada marca). Portanto, do ponto de vista teórico, a análise de desequilíbrio e o mapeamento de marcas moleculares são mais eficientes em F_2 , que equivale a um composto de segunda geração derivado de duas populações homozigóticas, uma vez que, para todo $r < 1/2$, a medida do desequilíbrio é maximizada.

Como em um composto derivado de duas populações a medida do desequilíbrio de fase gamética entre duas marcas é uma função da freqüência de gametas recombinantes, é possível o mapeamento de marcas moleculares com a genotipagem das populações genitoras. Usando as estimativas de máxima verossimilhança das freqüências das marcas nas populações cruzadas e da medida do desequilíbrio entre elas, a distância entre duas marcas é estimada por:

$$\hat{r} = \frac{1}{2} \left[1 - \frac{4\hat{\Delta}^{(-1)}}{(\hat{p}_a^1 - \hat{p}_a^2)(\hat{p}_b^1 - \hat{p}_b^2)} \right]$$

Infelizmente, vieses pequenos nas estimativas das freqüências de mesma marca nos genitores e na estimativa da medida do desequilíbrio tendem a promover viés elevado na estimativa da distância entre duas marcas. Por exemplo, se $p_a^1 = 0,7$, $p_a^2 = 0,8$, $p_b^1 = 0,5$ e $p_b^2 = 0,6$, então $\Delta^{(-1)} = 0,0045$. Se

não há viés nas estimativas dos parâmetros, $\hat{r} = 0,05$. Assumindo viés de 10% nas estimativas de cada parâmetro, $\hat{r} = 0,09$, uma estimativa com viés de 80%. Mesmo que os vieses nas estimativas dos parâmetros sejam reduzidos, o mapeamento em um composto derivado de duas populações tende a ser menos eficiente que em F_2 , uma vez que a medida do desequilíbrio pode ser muito pequena, inclusive para genes fisicamente muito próximos. Se as diferenças de freqüências de mesma marca nas populações genitoras são pequenas, a medida do desequilíbrio será de magnitude reduzida, dificultando declarar as duas marcas como ligadas (pertencentes ao mesmo grupo de desequilíbrio ou de ligação).

Portanto, do ponto de vista teórico, pode-se afirmar que a análise de desequilíbrio e o mapeamento de marcas moleculares tendem a ser mais eficientes quanto maiores forem as diferenças de freqüências de mesma marca nas populações cruzadas, e quanto menores os vieses nas estimativas das freqüências das marcas nos genitores e das medidas do desequilíbrio de fase gamética. É certo, então, que para o mesmo tamanho amostral o mapeamento é mais eficiente em F_2 do que em um composto derivado do cruzamento entre duas populações.

3.1.2- Identificação de QTL pela análise de marcas simples

Considere um caráter quantitativo, determinado por k genes bi-alélicos, sendo apenas um deles um *Quantitative Trait Locus*, e uma população em equilíbrio de *Hardy-Weinberg*. Os valores genotípicos médios dos indivíduos QQ, Qq e qq são:

$$G_{QQ} = E \left[M + 2q_i \alpha_i + \sum_{i'=1}^{k(i' \neq i)} A_{i'} + (-2q_i^2 d_i) + \sum_{i'=1}^{k(i' \neq i)} D_{i'} \right] = M + A_{QQ} + D_{QQ}$$

$$G_{Qq} = E \left[M + (q_i - p_i) \alpha_i + \sum_{i'=1}^{k(i' \neq i)} A_{i'} + 2p_i q_i d_i + \sum_{i'=1}^{k(i' \neq i)} D_{i'} \right] = M + A_{Qq} + D_{Qq}$$

$$G_{qq} = E \left[M + (-2p_i \alpha_i) + \sum_{i'=1}^{k(i' \neq i)} A_{i'} + (-2p_i^2 d_i) + \sum_{i'=1}^{k(i' \neq i)} D_{i'} \right] = M + A_{qq} + D_{qq}$$

sendo p e q frequências alélicas, α o efeito médio de uma substituição gênica, d o desvio de dominância, A o valor genético aditivo e D o valor genético devido à dominância (LYNCH e WALSH, 1998). Admitindo uma marca molecular bi-alélica de herança co-dominante em desequilíbrio da fase gamética em relação ao QTL, as médias dos grupos M_1M_1 , M_1M_2 e M_2M_2 são:

$$\begin{aligned}
 G_{M_1M_1} &= \frac{1}{p_b^2} \left(f_{22}^{(0)} G_{QQ} + f_{12}^{(0)} G_{Qq} + f_{02}^{(0)} G_{qq} \right) \\
 &= M + \left[\frac{\Delta^{(-1)}}{p_b q_b} \right] (2q_b \alpha) + \left[\frac{\Delta^{(-1)}}{p_b q_b} \right]^2 (-2q_b^2 d) \\
 &= M + A_{M_1M_1} + D_{M_1M_1} \\
 G_{M_1M_2} &= \frac{1}{2p_b q_b} \left[f_{21}^{(0)} G_{QQ} + (f_{11c}^{(0)} + f_{11r}^{(0)}) G_{Qq} + f_{01}^{(0)} G_{qq} \right] \\
 &= M + \left[\frac{\Delta^{(-1)}}{p_b q_b} \right] (q_b - p_b) \alpha + \left[\frac{\Delta^{(-1)}}{p_b q_b} \right]^2 (2p_b q_b d) \\
 &= M + A_{M_1M_2} + D_{M_1M_2} \\
 G_{M_2M_2} &= \frac{1}{q_b^2} \left(f_{20}^{(0)} G_{QQ} + f_{10}^{(0)} G_{Qq} + f_{00}^{(0)} G_{qq} \right) \\
 &= M + \left[\frac{\Delta^{(-1)}}{p_b q_b} \right] (-2p_b \alpha) + \left[\frac{\Delta^{(-1)}}{p_b q_b} \right]^2 (-2p_b^2 d) \\
 &= M + A_{M_2M_2} + D_{M_2M_2}
 \end{aligned}$$

em que p_b e q_b são as frequências alélicas, e A e D são os valores genéticos aditivo e devido à dominância, em relação ao loco do marcador do QTL. O efeito de substituição da marca M_2 pela marca M_1 é:

$$\alpha_M = \alpha_{M_1} - \alpha_{M_2} = \left[\frac{\Delta^{(-1)}}{p_b q_b} \right] (q_b \alpha) - \left[\frac{\Delta^{(-1)}}{p_b q_b} \right] (-p_b \alpha) = \frac{\Delta^{(-1)}}{p_b q_b} \alpha$$

As diferenças entre as médias dos grupos são:

$$G_{M_1M_1} - G_{M_1M_2} = \alpha_M - 2q_b d_M$$

$$G_{M_1M_1} - G_{M_2M_2} = 2\alpha_M - 2(q_b - p_b) d_M$$

$$G_{M_1M_2} - G_{M_2M_2} = c\alpha_M + 2p_b d_M$$

sendo $d_M = \left[\frac{\Delta^{(-1)}}{p_b q_b} \right]^2 d$ o desvio de dominância em relação ao loco do marcador.

Então, a significância de um ou mais desses contrastes (teste t) evidencia que o fragmento polimórfico é um marcador molecular de um QTL. Em caso de rejeição da hipótese de nulidade, pode-se testar a hipótese de que não há dominância no loco do marcador (teste t), a partir do contraste

$$G_{M_1M_2} - \frac{G_{M_1M_1} + G_{M_2M_2}}{2} = d_M$$

Alternativamente ao teste t, pode-se identificar QTL por análise de regressão (HALEY e KNOTT, 1992). Os resultados teóricos da análise de regressão são função de componentes da variância genotípica na população, em relação ao loco do marcador do QTL. Assim, a variância dos valores genotípicos médios dos indivíduos M_1M_1 , M_1M_2 e M_2M_2 , ou variância das médias dos grupos, é:

$$\begin{aligned} \sigma_{Ga(M)}^2 &= 2p_b q_b \left[\frac{\Delta^{(-1)}}{p_b q_b} \right]^2 \alpha^2 + 4p_b^2 q_b^2 \left[\frac{\Delta^{(-1)}}{p_b q_b} \right]^4 d^2 \\ &= \sigma_{A(M)}^2 + \sigma_{D(M)}^2 \end{aligned}$$

em que $\sigma_{A(M)}^2$ e $\sigma_{D(M)}^2$ são as variâncias aditiva e devido à dominância no loco do marcador.

As variâncias dos valores genotípicos dos indivíduos em cada grupo são:

$$\begin{aligned}
\sigma_{G(M_1M_1)}^2 &= \sigma_{G_1}^{2*} + 2 \left[p_a q_a - c^2 q_b^2 + c(q_a - p_a) q_b \right] \alpha^2 \\
&+ 4 \left[p_a^2 q_a^2 - c^4 q_b^4 + c^2 (q_a - p_a)^2 q_b^2 + 2c p_a q_a (q_a - p_a) q_b \right] d^2 \\
&- 8c q_b \left[p_a q_a - c^2 q_b^2 + c(q_a - p_a) q_b \right] \alpha d \\
\sigma_{G(M_1M_2)}^2 &= \sigma_{G_2}^{2*} + \left[2p_a q_a - c^2 (1 - 2p_b q_b) - c(q_a - p_a)(q_b - p_b) \right] \alpha^2 \\
&+ 4 \left[p_a^2 q_a^2 - c^4 p_b^2 q_b^2 - c^2 (q_a - p_a)^2 p_b q_b + c p_a q_a (q_a - p_a)(q_b - p_b) \right] d^2 \\
&- 4c \left[p_a q_a (q_b - p_b) + c^2 p_b q_b (q_b - p_b) - 2c(q_a - p_a) p_b q_b \right] \alpha d \\
\sigma_{G(M_2M_2)}^2 &= \sigma_{G_3}^{2*} + 2 \left[p_a q_a - c^2 p_b^2 - c(q_a - p_a) p_b \right] \alpha^2 \\
&+ 4 \left[p_a^2 q_a^2 - c^4 p_b^4 + c^2 (q_a - p_a)^2 p_b^2 - 2c p_a q_a (q_a - p_a) p_b \right] d^2 \\
&+ 8c p_b \left[p_a q_a - c^2 p_b^2 - c(q_a - p_a) p_b \right] \alpha d
\end{aligned}$$

em que σ_G^{2*} é a variância genotípica relativa aos $k - 1$ genes menores, e p_a e q_a são as freqüências alélicas do QTL.

A variância média dentro de grupo é:

$$\sigma_{Gw(M)}^2 = \sigma_G^{2*} + \sigma_{G(QTL)}^2 - \sigma_{Ga(M)}^2$$

Logo, a variância genotípica na população é:

$$\sigma_G^2 = \sigma_{Ga(M)}^2 + \sigma_{Gw(M)}^2$$

O modelo de regressão mais informativo em uma análise, visando a identificação de marcadores moleculares associados a QTL, é o que relaciona o fenótipo (P) do indivíduo com o número de cópias da marca (X). Visando adicionalmente inferir sobre dominância no loco do QTL, o modelo adequado é:

$$\text{Modelo (1)} \quad P_{M_i M_j} = \beta_0 + \beta_1 X + \beta_2 X^2 + \varepsilon,$$

sendo X, p. ex., o número de cópias da marca M_1 (0, 1 ou 2) e ε o erro ou resíduo. Alternativamente, pode-se ajustar o modelo (2)

$$P_{M_i M_i} = \beta_0 + a_M + \varepsilon$$

$$P_{M_i M_j} = \beta_0 + d_M + \varepsilon$$

$$P_{M_j M_j} = \beta_0 - a_M + \varepsilon$$

Esse modelo pressupõe que o fragmento molecular sob análise é marcador de um QTL e que, portanto, os desvios a e d relativos ao marcador são uma função dos desvios a (diferença entre o valor genotípico do homocigoto de maior expressão e a média dos valores genotípicos dos homocigotos) (LYNCH e WALSH, 1998) e d relativos ao QTL.

Em relação ao primeiro modelo, considerando os valores genotípicos médios relativos ao loco do QTL, tem-se:

$$\begin{bmatrix} G_{QQM_1M_1} \\ G_{QQM_1M_2} \\ G_{QQM_2M_2} \\ G_{QqM_1M_1} \\ G_{QqM_1M_2} \\ G_{QqM_2M_2} \\ G_{qqM_1M_1} \\ G_{qqM_1M_2} \\ G_{qqM_2M_2} \end{bmatrix} = \begin{bmatrix} 1 & 2 & 4 \\ 1 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 2 & 4 \\ 1 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 2 & 4 \\ 1 & 1 & 1 \\ 1 & 0 & 0 \end{bmatrix} \beta$$

A matriz de probabilidades dos genótipos é:

$$A = \begin{bmatrix} f_{22}^{(0)} & 0 & \dots & 0 \\ 0 & f_{21}^{(0)} & \dots & 0 \\ \dots & \dots & \ddots & \dots \\ 0 & 0 & \dots & f_{00}^{(0)} \end{bmatrix}$$

Empregando a técnica de polinômios ortogonais, demonstra-se que:

$$\beta' = [M \quad \alpha_M \quad -d_M]$$

$$R(\beta_1 | \beta_0, \beta_2) = \sigma_{A(M)}^2$$

$$R(\beta_2 | \beta_0, \beta_1) = \sigma_{D(M)}^2$$

Então, a constante da regressão é a média genotípica da população, o coeficiente angular é o efeito de substituição em relação à marca e o coeficiente do termo quadrático é o desvio de dominância em relação à marca. E mais, testar $H_0: \beta_1 = 0$ equivale a testar que o fragmento molecular não é marcador de um QTL. Em caso de rejeição dessa hipótese, testar $H_0: \beta_2 = 0$ equivale a testar que não há dominância no loco do QTL. Para estimação das variâncias aditiva e de dominância em relação ao marcador, basta ajustar o modelo linear simples com as médias dos três grupos. As estimativas são as somas de quadrados da regressão e do resíduo, respectivamente.

Em relação ao segundo modelo, considerando adicionalmente o termo quadrático, tem-se:

$$\begin{bmatrix} G_{QQM_1M_1} \\ G_{QQM_1M_2} \\ G_{QQM_2M_2} \\ G_{QqM_1M_1} \\ G_{QqM_1M_2} \\ G_{QqM_2M_2} \\ G_{qqM_1M_1} \\ G_{qqM_1M_2} \\ G_{qqM_2M_2} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & -1 & 1 \\ 1 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & -1 & 1 \\ 1 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & -1 & 1 \end{bmatrix} \beta$$

Empregando a técnica de polinômios ortogonais, demonstra-se que:

$$\beta' = \left[G_{M_1M_2} \quad \alpha_M - (q_b - p_b)d_M \quad -d_M \right]$$

$$R(\beta_2 | \beta_0, \beta_1) = \sigma_{D(M)}^2$$

Então, a constante da regressão é a média genotípica dos heterozigotos em relação à marca molecular, o coeficiente angular é uma função do efeito de substituição e do desvio de dominância em relação à marca e o coeficiente do termo quadrático é o desvio de dominância em relação à marca. Como no modelo anterior, testar $H_0: \beta_1 = 0$ equivale a testar que o fragmento molecular não é marcador de um QTL. Em caso de rejeição dessa hipótese, testar $H_0: \beta_2 = 0$ equivale a testar que não há dominância no loco do QTL.

3.2- Simulação dos Dados

3.2.1- Mapeamento de marcas moleculares

Para simulação dos genomas paramétricos, simulação dos genitores, dos compostos derivados do cruzamento entre duas populações parentais e análise dos dados, foi utilizado o aplicativo computacional *RealBreeding* (VIANA, 2007).

Foram simulados genomas com cinco grupos de ligação contendo 20 marcas bi-alélicas e co-dominantes por grupo, e com níveis de saturação alto e aleatório (Figuras 1 e 2). O genoma simulado com nível de saturação alto apresentou intervalos entre marcas adjacentes menores que 5 centiMorgan (cM), enquanto o genoma com nível de saturação aleatório apresentou intervalos entre marcas adjacentes, em cada grupo de ligação, variando entre 0 e 20 cM. Esses genomas foram utilizados para a geração das diferentes populações simuladas (compostos) e, posteriormente, como parâmetro na avaliação da recuperação dos mesmos.

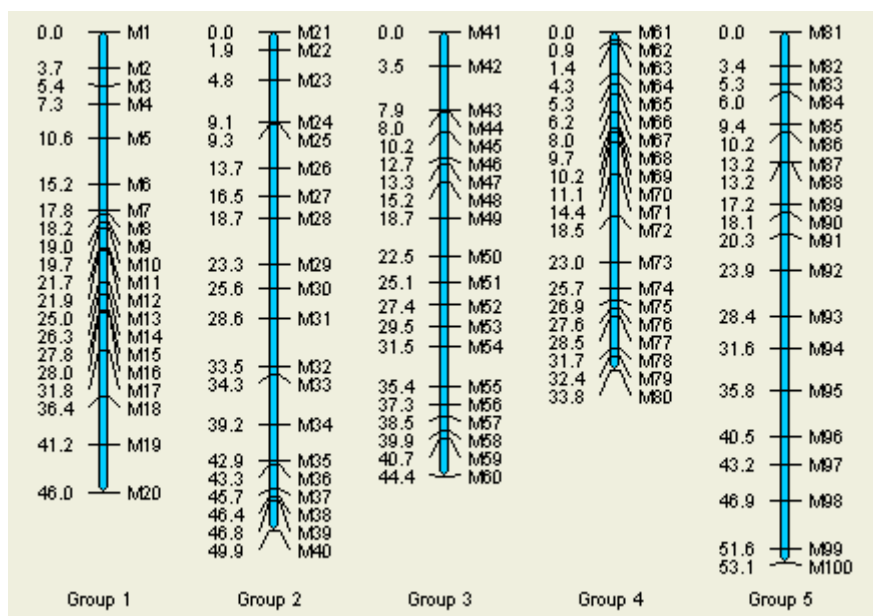


Figura 1 – Mapa genético com cinco grupos de ligação (GL), 20 marcas por grupo e nível de saturação alto (distância entre marcas adjacentes < 5 cM). Tamanho total do genoma igual a 227,2 cM.

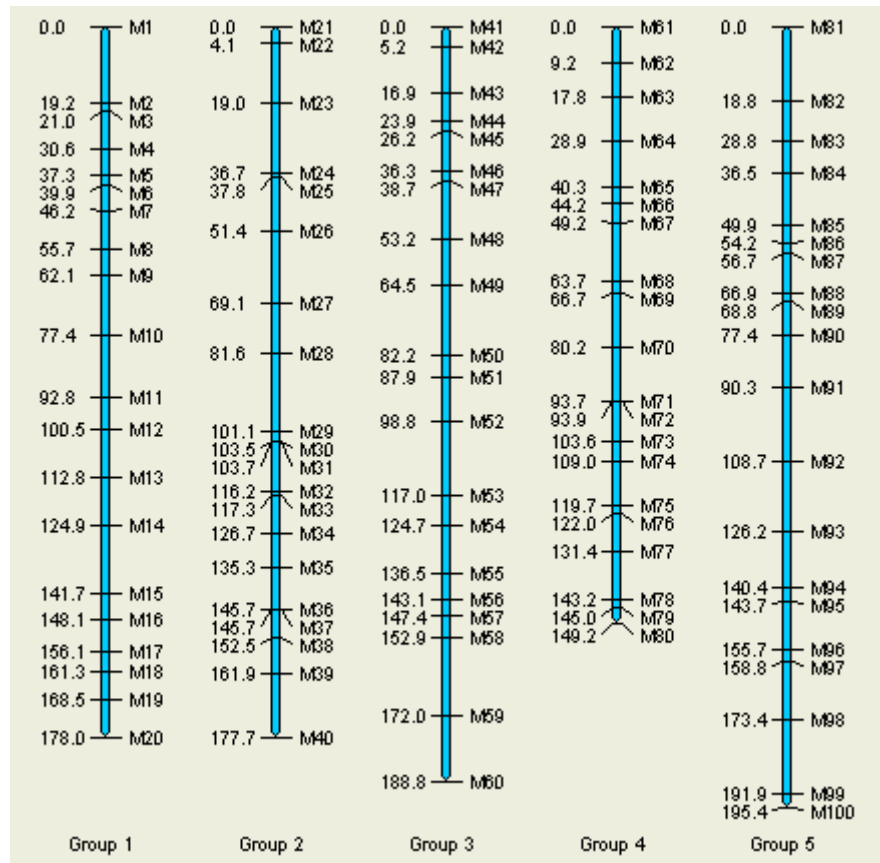


Figura 2 – Mapa genético com cinco grupos de ligação (GL), 20 marcas por grupo e nível de saturação aleatório (distância entre marcas adjacentes variando entre 0 e 20 cM). Tamanho total do genoma igual a 889,1 cM.

Foram simuladas amostras de compostos com 1.000 e 200 indivíduos, em equilíbrio de *Hardy-Weinberg*, derivados do cruzamento entre duas populações parentais com 200 indivíduos, também em equilíbrio de *Hardy-Weinberg*, considerando o modelo de recombinação de Haldane. Além dos parâmetros tamanho do composto e nível de saturação do genoma, na simulação das amostras dos compostos foram estabelecidos quatro níveis de diferença mínima de frequência de mesma marca entre as populações parentais (1, 0,9, 0,8 e 0,7). Como foram considerados dois níveis de saturação do genoma, dois tamanhos de amostra do composto e quatro valores de diferença mínima de frequência de marcas entre os genitores, foram gerados 16 compostos diferentes. Os compostos derivados de duas populações parentais com diferença mínima de frequência de mesma marca igual a 1,0 equivalem a uma população de mapeamento F_2 , na qual as frequências das marcas são conhecidas e iguais a 0,5. As amostras de cada composto foram obtidas com base em 50 repetições.

Com o intuito de verificar a razão de segregação de cada marca em todos os compostos gerados, foram aplicados testes do qui-quadrado (χ^2) a 5% de probabilidade. No processo de mapeamento foram utilizadas todas as marcas, mesmo aquelas que não segregaram de acordo com a proporção esperada de $p^2:2pq:q^2$ (AA:Aa:aa). O teste do qui-quadrado também foi utilizado para testar a hipótese de nulidade de que as marcas estavam em equilíbrio de *Hardy-Weinberg*, tanto nas populações parentais, como nos compostos. Este teste foi utilizado como uma medida para avaliar a eficiência do programa em gerar informações mais próximas possíveis de um sistema natural.

O próximo passo foi a determinação dos grupos de ligação. Na determinação dos grupos de ligação, para inferir se dois locos estavam ligados ou não, foram utilizados como critérios a frequência máxima de recombinação (r_{\max}) de 40% e LOD mínimo (LOD_{\min}) igual a 3. Após a formação dos grupos de ligação é preciso estabelecer a melhor ordem das marcas nesses grupos. Para isso, foi utilizado o método RCD (*Rapid Chain Delineation*), que se baseia na escolha da melhor ordem das marcas pela menor soma das recombinações adjacentes (SARF - *Sum of Adjacent Recombination Fractions*). Assim, considera-se a ordem original estabelecida pelo processo de agrupamento e realizam-se permutas entre dois marcadores vizinhos ou distantes, envolvendo três ou quatro marcadores. A ordem é alterada se, após a permuta, a soma das distâncias adjacentes for reduzida (SCHUSTER e CRUZ, 2004).

Por fim, foi feita a comparação dos genomas simulados com os genomas amostrais. Para melhor compreensão e efeito de comparação, o termo “genoma simulado” deve ser interpretado como paramétrico ou real. O termo “genoma amostral” refere-se aos genomas obtidos partir das populações simuladas (compostos), o qual apresentará distorção em razão do processo de estimação, amostragem, dentre outros.

A partir dos genomas amostrais foi estimado, em cada situação, o porcentual de simulações válidas, ou seja, o porcentual de simulações em que houve a recuperação dos cinco grupos de ligação do genoma paramétrico, a distância média entre marcas adjacentes, a variância das distâncias entre marcas adjacentes, o porcentual de marcas ordenadas corretamente, a correlação entre as distâncias das marcas adjacentes do genoma paramétrico

e as do genoma amostral, e o estresse. Essas variáveis foram obtidas por grupo de ligação e apresentadas com base na média das simulações em que houve a formação de cinco grupos de ligação no mapeamento genético de cada composto. Exceção feita apenas para as variáveis distância e variância média entre marcas adjacentes, em que foram utilizadas todas as simulações, independentemente da recuperação dos cinco grupos de ligação.

A correlação média fornece uma medida da mudança da distância entre as marcas adjacentes nos grupos de ligação do genoma amostral e a relação ao paramétrico. Valores de correlação iguais à unidade indicam que as distâncias entre as marcas adjacentes não foram alteradas em relação às distâncias paramétricas. Por outro lado, valores menores do que 1 indicam que as distâncias entre as marcas adjacentes no genoma amostral foram alteradas em relação às distâncias paramétricas.

O coeficiente de estresse (S) é comumente utilizado como medida de adequação da representação gráfica de medidas de dissimilaridade convertidas em escores relativos às variáveis X e Y em estudos de divergência genética (CRUZ e CARNEIRO, 2003). Sua utilização na análise de genomas foi proposta por Silva (2005), conforme apresentado a seguir:

$$S = 100 \cdot \sqrt{\frac{\sum_{k=1}^{m-1} (d_{ok} - d_k)^2}{\sum_{k=1}^{m-1} d_{ok}^2}}$$

em que S é o valor estimado do estresse, em porcentagem, para o grupo de ligação do genoma amostral; d_{ok} é a distância entre marcas adjacentes m_k e m_{k+1} no grupo de ligação do genoma amostral; d_k é a distância entre marcas adjacentes m_k e m_{k+1} no grupo de ligação do genoma paramétrico ($k= 1, \dots, m-1$); e m é o número de marcadores no grupo de ligação do genoma paramétrico e no grupo de ligação do genoma amostral. Neste contexto, se as distâncias entre os marcadores no genoma amostral mantiverem-se as mesmas, com relação ao genoma paramétrico, o valor estimado do estresse será zero, indicando a perfeita recuperação do genoma paramétrico.

3.2.2- Detecção de QTL

No estudo de identificação de QTL, também foi utilizado o aplicativo computacional *RealBreeding* para simulação dos genomas paramétricos, dos genitores, dos compostos derivados do cruzamento entre duas populações parentais e dos dados para a análise de QTL (VIANA, 2007).

Foram simulados genomas com dez grupos de ligação nos níveis de saturação alto e aleatório, devendo ser ressaltado que em cada genoma somente o grupo de ligação 1 foi simulado com 5 locos marcadores (Figuras 3 e 4). Nestes grupos de ligação também foram distribuídos 200 genes controladores de características quantitativas. Porém, somente no grupo de ligação 1 foram alocados um loco controlador de característica quantitativa (QTL), identificado como Q1, e 19 genes de efeito menor. Em cada um dos demais grupos de ligação foram alocados mais 20 genes de efeito menor.

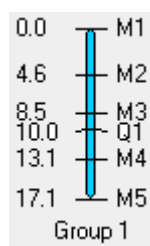


Figura 3 – Grupo de ligação (GL) com cinco marcas, nível de saturação alto (distância entre marcas adjacentes < 5 cM) e um QTL (Q1) flanqueado pelas marcas M3 e M4.

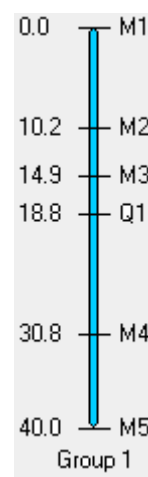


Figura 4 – Grupo de ligação (GL) com cinco marcas, nível de saturação aleatório (distância entre marcas adjacentes variando entre 0-20 cM) e um QTL (Q1) flanqueado pelas marcas M3 e M4

Para simulação das amostras dos compostos, foram adotados os mesmos critérios do mapeamento de marcas moleculares, como nível de saturação do genoma, tamanho das populações parentais, tamanho de amostra do composto e número de repetições, porém com apenas três valores de diferença mínima de frequência de mesma marca entre os genitores (1, 0,9 e 0,8). Como foram considerados dois níveis de saturação do genoma, dois

tamanhos amostra do composto e três diferenças mínimas de freqüência de mesma marca entre as populações parentais (2 x 2 x 3), foram simulados 12 compostos diferentes.

O passo seguinte foi a obtenção dos dados para análise de QTL. Para isso, foram idealizadas três características quantitativas, cada uma governada por vários locos gênicos, porém com apenas um QTL (Q_1) associado a cada característica, como descrito anteriormente, e identificado nas populações de compostos simuladas. As características idealizadas foram produção (PROD), com direção de dominância positiva (gene favorável dominante); capacidade de expansão (CE), com direção de dominância bidirecional; e crescimento líquido relativo (CLR), com direção de dominância negativa (gene favorável recessivo). Os valores genotípicos máximo e mínimo para PROD, CE e CLR foram de 150 e 10 g/espiga, 50 e 5 ml/g e de 130 e 0%, respectivamente. Essas características são normalmente avaliadas em programas de melhoramento da cultura do milho-pipoca. Ainda em relação aos QTL, dois graus de dominância foram considerados para cada característica, em uma primeira condição sob dominância completa e em uma segunda sob dominância parcial. Também foram considerados três níveis de herdabilidade (90, 60 e 30%). Mediante a informação dos valores genotípicos referentes a cada característica, do grau e direção de dominância e das herdabilidades, foram obtidos os percentuais de variação fenotípica explicada pelo QTL em questão.

Neste trabalho foi empregado o método de mapeamento da marca simples, com base em análise de regressão. Assim, para avaliar a associação entre as marcas moleculares e os QTL, e os efeitos de substituição de marca e de dominância no loco do marcador, foram utilizados os modelos de regressão linear e quadrático, pelo procedimento *PROC REG* do programa SAS (2002). As hipóteses de nulidade $H_0: \beta_1 = 0$ (o marcador não está associado ao QTL) e $H_0: \beta_2 = 0$ (não há dominância no loco do marcador) foram testadas a 5% de probabilidade, pela estatística “t”. O poder de detecção do QTL foi estabelecido com base no percentual de testes significativos a 5%, pelo modelo de regressão linear, ou seja, percentual de repetições em que houve associação significativa entre os marcadores e o QTL associado a cada uma das características, nas 50 repetições. Procedeu-se da mesma forma para o teste de dominância nos locos dos marcadores, porém com base no modelo

quadrático. Os valores estimados dos efeitos de substituição de marca e dos desvios de dominância, obtidos pelos modelos linear e quadrático, respectivamente, também foram apresentados com base na média das 50 repetições.

4- RESULTADOS E DISCUSSÃO

4.1- Mapeamento de marcas moleculares

No procedimento de simulação, para a construção de mapas genéticos e análise de QTL, é importante que as marcas simuladas estejam em equilíbrio de *Hardy-Weinberg* (EHW). Isto para que se tenham duas populações parentais em equilíbrio de *Hardy-Weinberg* e que gerem compostos também em equilíbrio de Hardy-Weinberg. Esta medida é uma forma de avaliar a eficiência do programa na simulação de um sistema natural. Assim, neste trabalho foi utilizado o teste do qui-quadrado, com nível de significância alfa 5% (erro tipo I), para testar a hipótese de nulidade (marcas em EHW). Mesmo as marcas que apresentaram significância acima de 5% foram mantidas, o que não interferiu negativamente na análise. Em geral, o percentual médio de testes do qui-quadrado significativos foi abaixo de 5%, como esperado, o que indica adequação do processo de simulação (Tabela 1).

A simulação tem sido importante em estudos genéticos sob vários contextos, incluindo estudos de populações, do indivíduo ou do próprio genoma (CRUZ, 2001). No entanto, para a garantia da eficiência dessa técnica, o pesquisador deve lançar mão de processos de validação. Essa validação

consiste em fazer o sistema simulado operar nas condições do sistema real e verificar, por meio de testes estatísticos, se os resultados obtidos com a simulação condizem com os observados no sistema real (FERREIRA, 2001).

Silva (2005), por meio de simulação, buscando determinar o número adequado de indivíduos a ser utilizado no mapeamento de RIL (*Recombinant Inbred Line*), assim como o efeito do nível saturação do genoma por marcas moleculares, concluiu que para obtenção de mapas com boa confiabilidade devem ser utilizados no mínimo 200 indivíduos no caso de saturação alta do genoma, ou seja, com distância média entre marcas adjacentes igual a 5 cM.

Tabela 1 – Porcentual médio de testes do qui-quadrado, significativos a 5% para equilíbrio de *Hardy-Weinberg*, para as marcas moleculares nas populações parentais e nos compostos, em função do nível de saturação do mapa, do tamanho do composto e da diferença mínima de frequência de mesma marca entre os genitores

Nível de saturação do mapa	Tamanho do composto	Populações	Diferença mínima de frequência das marcas			
			1,0	0,9	0,8	0,7
Alto	1000	P ₁	0,00	3,84	4,24	3,76
		P ₂	0,00	3,34	4,46	4,66
		Composto	5,34	5,88	4,78	4,42
	200	P ₁	0,00	4,14	3,94	3,76
		P ₂	0,00	4,30	4,40	4,56
		Composto	5,26	5,04	4,72	4,78
Aleatório	1000	P ₁	0,00	4,22	4,06	4,10
		P ₂	0,00	3,92	3,82	4,10
		Composto	4,94	5,32	4,18	5,40
	200	P ₁	0,00	3,50	4,46	4,30
		P ₂	0,00	4,34	4,24	4,30
		Composto	5,10	4,78	4,98	4,98

P₁ e P₂ = populações parentais; e composto = compostos derivados do cruzamento entre duas populações (segunda geração).

No processo de mapeamento com base em desequilíbrio de fase gamética, utilizando populações não-endogâmicas e não-estruturadas em famílias, verificou-se que os grupos de ligação foram recuperados de forma satisfatória, com exceção apenas de três situações (Tabela 2). Nos compostos em que não se recuperou os cinco grupos de ligação em todas as repetições, foram observadas tanto a fusão quanto a fragmentação de grupos. Nestas situações o número de grupos de ligação recuperado no processo de mapeamento variou entre o máximo de 23 e o mínimo de três grupos.

Analisando as estimativas das distâncias médias entre as marcas adjacentes, pode-se observar que à medida que se diminuiu a diferença de

freqüência de mesma marca entre as populações parentais ou genitores maiores foram os vieses nessa estimativa (Tabela 2). É importante ressaltar que valores de distância média entre marcas adjacentes mais próximos dos valores paramétricos demonstram maior precisão na recuperação do genoma paramétrico.

Tabela 2 – Parâmetros de eficiência na construção de mapas genéticos obtidos a partir de populações naturais, em função do nível de saturação do mapa, do tamanho do composto e da diferença mínima de freqüência de mesma marca entre os genitores

Nível de saturação do mapa	Tamanho do composto	Genomas	Estatísticas	Diferença mínima de freqüência das marcas			
				1,0	0,9	0,8	0,7
Alto	1000	Paramétrico	Distância média	2,39	2,39	2,39	2,39
			Variância média	2,09	2,09	2,09	2,09
		Amostral	Sim. válidas (%)	100	100	100	14
			Distância média.	2,42	11,36	19,33	23,82
			Variância média	2,21	34,08	80,50	133,24
			Marcas ord. cor. (%)	99,36	94,08	94,40	93,71
	Correlação.	0,97	0,29	0,22	0,15		
	Estresse	0,19	179,37	589,36	1003,19		
	200	Paramétrico	Sim. válidas (%)	100	96	84	0
			Distância média.	2,53	11,46	20,02	21,94
		Amostral	Variância média	3,63	40,84	95,66	99,84
			Marcas ord. cor. (%)	95,15	89,58	86,30	-
			Correlação.	0,88	0,23	0,18	-
			Estresse	0,10	189,62	643,84	-
Aleatório	1000	Paramétrico	Distância média	9,36	9,36	9,36	9,36
			Variância média	29,54	29,54	29,54	29,54
		Sim. válidas (%)	100	100	100	100	
	Amostral	Distância média.	9,38	11,38	14,53	19,85	
		Variância média	29,98	33,86	42,49	66,44	
		Marcas ord. cor. (%)	97,52	99,32	97,44	93,95	
		Correlação.	0,99	0,88	0,74	0,52	
	Estresse	0,20	4,96	19,25	62,81		
	200	Paramétrico	Sim. válidas (%)	100	94	88	0
			Distância média.	9,61	11,86	15,70	17,07
Amostral		Variância média	34,69	41,02	55,31	65,22	
		Marcas ord. cor. (%)	96,36	97,43	91,51	-	
		Correlação.	0,96	0,82	0,63	-	
Estresse	1,01	7,33	28,16	-			

. Distância média = distância média entre as marcas adjacentes;

. Variância média = é a variância média da distância entre marcas adjacentes;

. Sim. válidas (%) = percentual de simulações válidas, ou seja, recuperação dos cinco grupos de ligação;

. Marcas ord. cor. (%) = percentual médio de marcas ordenadas corretamente;

. Correlação = correlação média das distâncias entre marcas adjacentes; e

. Estresse = estresse médio.

A variância média entre marcas adjacentes e o coeficiente de estresse também foram maiores à medida que houve diferença mínima de frequência de mesma marca entre as populações parentais, indicando maior viés na recuperação do genoma paramétrico. Desta forma, compostos derivados do cruzamento entre genitores com diferença de frequência de mesma marca igual a 1, compostos estes equivalentes a uma F_2 , foram os mais eficientes na recuperação do genoma paramétrico, pois praticamente não apresentaram vieses nessas estimativas. Também foi verificado que compostos obtidos com nível de saturação aleatório do mapa apresentaram menores vieses dessas estimativas em relação aos compostos com nível alto de saturação do mapa, independentemente do número de indivíduos genotipados. Por exemplo, nos compostos com nível alto de saturação do mapa, obtidos de genitores com diferença de frequência de mesma marca igual a 0,9, a razão entre distância média estimada e paramétrica e em relação ao composto equivalente a uma F_2 foi de aproximadamente 4,6 vezes, independentemente do tamanho do composto. Para esta mesma situação, porém com nível aleatório de saturação do mapa, esta razão foi de apenas 1,2 vez. Assim, pode-se afirmar que o mapeamento de marcas moleculares por meio de compostos derivados do cruzamento entre duas populações, com diferença de frequência de mesma marca menor ou igual a 0,9, é mais eficiente quando nível de saturação do mapa for aleatório.

O percentual de marcas ordenadas corretamente e a correlação média entre distâncias paramétricas e amostral foram analisados somente quando, em pelo menos uma das 50 repetições, houve a recuperação dos cinco grupos de ligação (Tabela 2). Em relação ao percentual de marcas ordenadas corretamente, o processo de mapeamento mostrou-se eficiente, pois o percentual mínimo de marcas ordenadas corretamente foi superior a 86%. Valores de correlação mais baixos foram observados nos compostos obtidos de genitores com menores diferenças de frequência de mesma marca e com nível alto de saturação do mapa, independente do tamanho amostral. Estes baixos valores de correlação indicam que, apesar de ordenadas corretamente, essas marcas apresentam um viés na estimativa das distâncias entre as elas, conforme já ressaltado.

4.2- Detecção de QTL

Na análise de QTL, de forma semelhante ao realizado para a construção dos mapas genéticos, também foi feito o teste do qui-quadrado para equilíbrio de *Hardy-Weinberg* das marcas moleculares e dos genes nas populações parentais e compostos. Em geral o percentual médio de testes do qui-quadrado significativos foi abaixo de 5% como esperado, indicando adequação do processo de simulação (Tabela 3).

Tabela 3 – Percentual médio de testes do qui-quadrado, significativos a 5% para equilíbrio de *Hardy-Weinberg*, para as marcas moleculares nas populações parentais e nos compostos, em função do nível de saturação do mapa, do tamanho do composto e da diferença de frequência de mesma marca entre os genitores

Nível de saturação do mapa	Tamanho do composto	Populações	Diferença mínima de frequência das marcas		
			1,0	0,9	0,8
Alto	1.000	P ₁	0,00	4,05	3,97
		P ₂	0,00	4,03	4,07
		Composto	4,78	4,84	5,21
	200	P ₁	0,00	3,50	4,18
		P ₂	0,00	3,54	4,16
		Composto	5,21	4,85	5,26
Aleatório	1.000	P ₁	0,00	4,26	3,73
		P ₂	0,00	4,22	4,30
		Composto	5,03	4,85	5,47
	200	P ₁	0,00	3,73	4,10
		P ₂	0,00	4,44	4,00
		Composto	5,63	5,19	4,52

P₁ e P₂ = populações parentais; e composto = compostos derivados do cruzamento entre duas populações (segunda geração).

4.2.1- Avaliação do poder de detecção do qtl

As porcentagens da variância fenotípica explicada pelo QTL das características produção, capacidade de expansão e crescimento líquido relativo apresentaram correspondência linear com os valores paramétricos de herdabilidade, sendo da ordem de um terço do valor de cada herdabilidade (Tabelas 4, 5 e 6).

Em relação ao poder de detecção de QTL, é necessário ressaltar que o resultado foi considerado satisfatório quando no mínimo 70% das repetições apresentaram testes significativos a 5% de probabilidade para os marcadores, ou seja, quando em pelo menos 70% das repetições o fragmento molecular foi

declarado marcador do QTL. Assim, na análise de detecção de QTL foi verificado, de modo geral, poder de detecção satisfatório associado a compostos derivados do cruzamento entre genitores com diferença mínima de frequência de mesma marca igual ou superior a 0,8, independentemente do caráter idealizado (Tabelas 4, 5 e 6). Mesmo nos compostos em que foi constatado menor poder de detecção pelas marcas na região do QTL, ao se avaliar apenas as marcas flanqueadoras, ainda verificou-se poder de detecção satisfatório, com exceção apenas para caráter crescimento líquido relativo com dominância completa negativa. Nessas situações de menor poder de detecção do QTL, também foi observado que à medida que se reduziu o valor da herdabilidade o poder de detecção também diminuiu. Esse fato pode ser justificado pela redução na variação fenotípica da característica explicada pelo QTL, sendo esta da ordem de um terço de cada herdabilidade.

Ainda em relação às situações em que o poder de detecção do QTL foi inferior a 70%, esse fato foi associado apenas compostos de tamanho amostral 200 e derivados do cruzamento entre genitores com diferença de frequência de mesma marca menor ou igual a 0,9. De modo geral, o poder de detecção de QTL não foi influenciado pelos níveis de saturação do mapa, grau e direção de dominância. Também é importante salientar que os resultados aqui obtidos são válidos para QTL que expliquem entre 10 e 30% da variação fenotípica do caráter estudado (Tabelas 4, 5 e 6).

Tabela 4 – Poder de detecção do QTL em porcentagem de testes significativos a 5% no modelo de regressão linear, para cada marca molecular, utilizando o método da marca simples, em função da característica produção (g/planta) com direção de dominância positiva, nível de saturação do mapa, tamanho do composto, grau de dominância, herdabilidade e diferença de frequência de mesma marca entre os genitores

Nível de saturação do mapa	Tamanho do composto	Grau de dominância	h ²	Diferença mínima de frequência das marcas																	
				1,0					0,9					0,8							
				VG	M1	M2	M3 *	M4	M5	VG	M1	M2	M3 *	M4	M5	VG	M1	M2	M3 *	M4	M5
Alto	1.000	1	90	30,4	100	100	100	100	100	29,5	100	100	100	100	100	30,2	100	100	100	100	100
			60	20,2	100	100	100	100	100	19,7	100	100	100	100	100	20,1	100	100	100	100	100
			30	10,1	100	100	100	100	100	9,9	100	100	100	100	100	10,1	100	100	100	100	100
		0-1	90	30,3	100	100	100	100	100	30,4	100	100	100	100	100	29,8	100	100	100	100	100
			60	19,9	100	100	100	100	100	20,7	100	100	100	100	100	29,6	100	100	100	100	100
			30	10,7	100	100	100	100	100	10,1	100	100	100	100	100	11,0	98	94	98	98	96
	200	1	90	30,1	100	100	100	100	100	29,7	44	64	90	76	42	29,8	100	100	100	100	100
			60	20,0	100	100	100	100	100	19,8	28	52	80	64	24	19,8	98	98	100	100	92
			30	10,0	100	100	100	100	100	9,9	12	32	54	46	12	9,9	66	74	96	100	78
		0-1	90	29,8	100	100	100	100	100	29,8	94	100	100	100	90	30,2	96	96	100	100	90
			60	19,1	100	100	100	100	100	20,7	84	98	100	98	88	20,1	60	62	100	92	50
			30	9,7	100	100	100	100	100	11,3	52	76	92	80	58	9,7	42	64	84	84	46
Aleatório	1.000	1	90	29,8	100	100	100	100	30,1	100	100	100	100	100	29,8	100	100	100	98	40	
			60	19,9	100	100	100	100	100	20,0	100	100	100	100	100	19,9	100	100	100	90	42
			30	10,0	100	100	100	100	100	10,0	100	100	100	100	100	10,0	78	84	96	52	12
		0-1	90	29,9	100	100	100	100	100	30,0	100	100	100	100	100	30,5	100	100	100	100	60
			60	21,5	100	100	100	100	100	16,7	100	100	100	100	100	20,0	100	100	100	100	52
			30	8,7	100	100	100	100	100	9,6	98	100	100	100	100	10,0	94	100	100	70	16
	200	1	90	29,5	100	100	100	100	100	30,5	100	100	100	100	98	30,2	88	100	100	96	58
			60	19,7	100	100	100	100	100	20,3	100	100	100	100	94	20,1	60	96	100	90	48
			30	9,9	98	100	100	98	98	10,1	76	94	100	88	48	10,0	30	60	88	54	26
		0-1	90	30,8	100	100	100	100	100	30,2	98	100	100	100	94	30,0	16	48	88	62	20
			60	20,5	100	100	100	100	98	20,6	80	98	100	96	72	18,4	22	42	88	46	10
			30	9,2	98	100	100	100	98	9,8	90	100	100	98	84	10,4	10	28	36	10	10

M₁ a M₅ = marcas moleculares; grau de dominância 1 = dominância completa e 0-1 = dominância parcial; h² = herdabilidade em porcentagem; VG= percentual da variância fenotípica explicada pelo QTL; e * = localização do QTL.

Tabela 5 – Poder de detecção do QTL em porcentagem de testes significativos a 5% no modelo de regressão linear, para cada marca molecular, utilizando o método da marca simples, em função da característica capacidade de expansão (mg/l) com direção de dominância bidirecional, nível de saturação do mapa, tamanho do composto, grau de dominância, herdabilidade e diferença de frequências de mesma marca entre os genitores

Nível de saturação do mapa	Tamanho do composto	Grau de dominância	h ²	Diferença mínima de frequência das marcas																			
				1,0					0,9					0,8									
				VG	M1	M2	M3 *	M4	M5	VG	M1	M2	M3 *	M4	M5	VG	M1	M2	M3 *	M4	M5		
Alto	1.000	1	90	30,4	100	100	100	100	100	29,6	100	100	100	100	100	29,8	100	100	100	100	100		
			60	20,2	100	100	100	100	100	19,5	100	100	100	100	100	15,7	100	100	100	100	100		
			30	10,1	100	100	100	100	100	8,7	100	100	100	100	100	10,1	100	100	100	100	100		
		0-1	90	29,8	100	100	100	100	100	30,0	100	100	100	100	100	29,5	100	100	100	100	100		
			60	16,3	100	100	100	100	100	21,3	100	100	100	100	100	20,7	100	100	100	100	100		
			30	9,2	100	100	100	100	100	11,0	100	100	100	100	100	10,0	100	100	100	100	100		
	200	1	90	30,2	100	100	100	100	100	30,4	94	100	100	98	84	30,3	60	66	90	92	78		
			60	20,1	100	100	100	100	100	20,0	70	90	100	92	72	20,0	22	34	56	60	30		
			30	10,0	100	100	100	100	100	11,3	10	12	30	16	10	12,7	70	78	100	100	78		
		0-1	90	29,8	100	100	100	100	100	30,4	90	98	100	100	82	30,4	100	100	100	100	100		
			60	19,4	100	100	100	100	100	19,4	56	84	98	88	70	18,7	68	70	96	98	66		
			30	10,0	100	100	100	100	100	9,1	36	50	74	66	34	9,1	64	64	86	94	46		
Aleatório	1.000	1	90	29,8	100	100	100	100	100	30,0	100	100	100	100	100	30,4	100	100	100	100	92		
			60	19,9	100	100	100	100	100	20,0	100	100	100	100	100	21,0	100	100	100	100	92		
			30	10,0	100	100	100	100	100	9,4	100	100	100	100	100	10,1	78	82	90	50	18		
		0-1	90	30,2	100	100	100	100	100	30,0	100	100	100	100	100	30,4	100	100	100	100	98		
			60	19,0	100	100	100	100	100	17,1	100	100	100	100	100	20,9	100	100	100	94	54		
			30	10,0	100	100	100	100	100	10,5	100	100	100	100	100	9,2	100	100	100	88	48		
	200	1	90	29,9	100	100	100	100	100	30,4	76	94	98	82	52	30,6	80	100	100	94	66		
			60	20,0	100	100	100	100	98	20,2	70	90	92	72	50	20,3	64	80	100	90	42		
			30	10,0	94	100	100	98	86	10,8	82	96	100	94	58	10,8	4	2	14	18	0		
		0-1	90	29,6	100	100	100	100	100	30,4	100	100	100	100	98	30,2	30	70	98	88	46		
			60	19,7	100	100	100	100	100	20,0	92	100	100	98	68	17,6	36	64	94	70	34		
			30	8,8	98	100	100	100	94	9,7	40	90	98	82	50	10,0	36	52	88	58	18		

M₁ a M₅ = marcas moleculares; grau de dominância 1 = dominância completa e 0-1 = dominância parcial; h² = herdabilidade em porcentagem; VG= percentual da variância fenotípica explicada pelo QTL; e * = localização do QTL.

Tabela 6 – Poder de detecção do QTL em porcentagem de testes significativos a 5% no modelo de regressão linear, para cada marca molecular, utilizando o método da marca simples, em função da característica crescimento líquido relativo (%) com direção de dominância negativa, nível de saturação do mapa, tamanho do composto, grau de dominância, herdabilidade e diferença de frequência de mesma marca entre os genitores

Nível de saturação do mapa	Tamanho do composto	Grau de dominância	h ²	Diferença mínima de frequência das marcas																			
				1,0					0,9					0,8									
				VG	M1	M2	M3 *	M4	M5	VG	M1	M2	M3 *	M4	M5	VG	M1	M2	M3 *	M4	M5		
Alto	1.000	1	90	30,4	100	100	100	100	100	30,0	100	100	100	100	100	30,3	100	100	100	100	100		
			60	20,2	100	100	100	100	100	20,0	100	100	100	100	100	20,2	100	100	100	100	100		
			30	10,1	100	100	100	100	100	10,0	100	100	100	100	100	10,1	100	100	100	100	100		
		0-1	90	29,8	100	100	100	100	100	30,5	100	100	100	100	100	29,8	100	100	100	100	100		
			60	16,9	100	100	100	100	100	21,7	100	100	100	100	100	20,6	100	100	100	100	100		
			30	9,5	100	100	100	100	100	12,2	100	100	100	100	100	11,6	100	100	100	100	100		
	200	1	90	30,2	100	100	100	100	100	29,9	10	24	44	38	10	30,4	100	100	100	100	100		
			60	20,1	100	100	100	100	100	19,9	6	14	36	22	12	20,3	94	100	100	100	96		
			30	10,0	100	100	100	98	100	10,0	8	8	18	6	4	10,1	70	76	96	94	58		
		0-1	90	29,7	100	100	100	100	100	29,7	40	66	90	82	36	30,0	100	100	100	100	100		
			60	19,8	100	100	100	100	100	17,3	42	58	86	76	36	20,0	100	98	100	100	98		
			30	8,8	100	100	100	100	100	10,2	8	24	42	22	18	9,1	46	60	94	84	56		
Aleatório	1.000	1	90	29,8	100	100	100	100	30,8	100	100	100	100	100	29,9	100	100	100	100	100			
			60	19,9	100	100	100	100	100	20,6	100	100	100	100	100	19,9	100	100	100	100	98		
			30	10,0	100	100	100	100	100	10,2	100	100	100	100	100	10,0	100	100	100	100	70		
		0-1	90	30,2	100	100	100	100	100	30,8	100	100	100	100	100	30,0	100	100	100	100	100		
			60	19,0	100	100	100	100	100	24,4	100	100	100	100	100	20,1	100	100	100	100	96		
			30	11,2	100	100	100	100	100	10,6	100	100	100	100	100	10,5	100	100	100	100	76		
	200	1	90	29,6	100	100	100	100	100	30,2	100	100	100	100	96	30,6	86	98	100	98	56		
			60	24,3	100	100	100	100	100	20,1	98	100	100	100	90	20,4	78	90	100	88	50		
			30	12,1	96	98	100	100	94	10,0	78	98	100	94	62	10,2	42	58	96	54	18		
		0-1	90	30,0	100	100	100	100	100	30,2	100	100	100	100	96	30,3	86	100	100	98	64		
			60	19,0	100	100	100	100	100	19,3	100	100	100	100	86	17,9	58	76	100	82	44		
			30	9,3	96	100	100	100	98	10,0	80	94	98	96	58	10,0	42	66	94	56	22		

M₁ a M₅ = marcas moleculares; grau de dominância 1 = dominância completa e 0-1 = dominância parcial; h² = herdabilidade em porcentagem; VG= percentual da variância fenotípica explicada pelo QTL; e * = localização do QTL.

4.2.2- Teste de dominância no loco do marcador

Na avaliação do teste de dominância com base no termo quadrático do modelo de regressão, o poder do teste em revelar dominância nos locos marcadores foi considerado eficiente quando, em pelo menos 70% das repetições, o teste foi significativo a 5% de probabilidade.

Em relação às três características idealizadas foi detectada dominância de forma satisfatória, pelas marcas na região do QTL, somente nos compostos derivados do cruzamento entre genitores com diferença mínima de frequência de mesma marca maior ou igual a 0,9 e com amostra de 1.000 indivíduos genotipados (Tabelas 7, 8 e 9). Algumas exceções, principalmente nos compostos com diferença de 0,9, foram observadas à medida que se reduziu a herdabilidade, entretanto mesmo com menor herdabilidade as marcas flaqueadoras ainda apresentaram poder de detecção superior a 70%.

Nos compostos equivalentes a uma F_2 , porém com 200 indivíduos genotipados, foi detectada dominância de forma satisfatória apenas quando dominância completa e nos locos marcadores mais próximos do QTL (Tabelas 7, 8 e 9). Exceção para a característica crescimento líquido relativo, quando houve nível alto de saturação do mapa, em que todos marcadores detectaram dominância de forma satisfatória quando dominância completa e quando dominância parcial, nas herdabilidades de 90 e 60% (Tabela 9). Para as três características, em todos os compostos em que foi detectada dominância em todos os locos marcadores associados ao QTL de forma satisfatória, o poder do teste não foi influenciado pelo nível de saturação do mapa.

No entanto, a detecção de dominância nos locos marcadores foi dependente do número de indivíduos genotipados e da diferença mínima de frequência de mesma marca entre os genitores, sendo mais eficiente em compostos com tamanho de amostra igual a 1.000 e diferença de frequência de mesma marca entre genitores maior ou igual 1 (Tabelas 7, 8 e 9).

Tabela 7 – Teste para dominância no loco do QTL em porcentagem de testes significativos a 5% no modelo de regressão quadrático, para cada marca molecular, utilizando o método da marca simples, em função da característica produção (g/planta) com direção de dominância positiva, nível de saturação do mapa, tamanho do composto, grau de dominância, herdabilidade e diferença de frequência de mesma marca entre os genitores

Nível de saturação do mapa	Tamanho do composto	Grau de dominância	h ²	Diferença mínima de frequência das marcas														
				1,0					0,9					0,8				
				M1	M2	M3 *	M4	M5	M1	M2	M3 *	M4	M5	M1	M2	M3 *	M4	M5
Alto	1.000	1	90	100	100	100	100	100	78	100	100	100	100	74	64	92	84	66
			60	100	100	100	100	100	60	86	100	100	94	60	42	78	58	46
			30	100	100	100	100	100	34	64	92	96	64	12	30	42	38	20
		0-1	90	100	100	100	100	100	66	80	98	88	76	82	58	84	74	60
			60	98	100	100	100	100	56	82	98	96	80	10	6	10	8	10
			30	96	96	100	100	96	14	36	56	50	28	40	40	56	52	38
	200	1	90	90	98	100	100	96	16	28	64	36	6	12	22	30	32	20
			60	78	90	92	94	84	18	30	50	20	4	8	8	20	16	16
			30	46	58	62	60	48	6	14	18	16	6	12	18	14	12	12
		0-1	90	66	84	92	90	74	4	6	12	8	4	28	20	52	40	30
			60	18	32	32	30	16	22	22	42	38	18	16	8	22	26	20
			30	14	20	32	28	20	28	24	56	20	8	12	2	6	8	4
Aleatório	1.000	1	90	100	100	100	100	100	84	100	100	100	50	28	58	86	18	14
			60	100	100	100	100	100	56	98	100	94	36	16	52	68	16	8
			30	98	100	100	100	98	36	86	80	72	26	14	16	32	8	12
		0-1	90	100	100	100	100	100	82	100	100	100	46	26	44	66	18	14
			60	100	100	100	100	100	36	72	56	50	18	14	12	12	10	4
			30	24	36	60	46	36	22	54	66	62	12	12	6	8	14	4
	200	1	90	78	98	100	94	54	18	18	20	4	0	18	28	38	8	4
			60	58	90	98	74	36	8	10	14	6	2	10	22	32	6	18
			30	24	64	68	36	24	0	4	10	16	6	8	12	16	8	4
		0-1	90	62	94	100	90	58	10	24	46	24	8	24	22	58	22	4
			60	64	88	94	74	52	6	24	44	12	12	12	14	18	8	8
			30	14	26	44	24	18	8	6	6	8	2	10	6	20	4	2

M₁ a M₅ = marcas moleculares; grau de dominância 1 = dominância completa e 0-1 = dominância parcial; h² = herdabilidade em porcentagem; e * = localização do QTL.

Tabela 8 – Teste para dominância no loco do QTL em porcentagem de testes significativos a 5% no modelo de regressão quadrático, para cada marca molecular, utilizando o método da marca simples, em função da característica capacidade de expansão (mg/l) com direção de dominância bidirecional, nível de saturação do mapa, tamanho do composto, grau de dominância, herdabilidade e diferença de frequência de mesma marca entre os genitores

Nível de saturação do mapa	Tamanho do composto	Grau de dominância	h ²	Diferença mínima de frequência das marcas														
				1,0					0,9					0,8				
				M1	M2	M3 *	M4	M5	M1	M2	M3 *	M4	M5	M1	M2	M3 *	M4	M5
Alto	1.000	1	90	100	100	100	100	100	84	100	100	100	100	72	66	90	88	76
			60	100	100	100	100	100	72	98	100	100	100	46	30	60	64	58
			30	92	100	100	100	96	20	60	100	100	82	10	12	28	28	22
		0-1	90	100	100	100	100	100	4	16	64	76	66	8	10	20	20	12
			60	10	10	8	4	0	18	54	92	98	88	44	44	82	82	60
			30	62	78	94	88	76	8	36	60	64	38	6	4	12	8	6
	200	1	90	92	100	100	100	98	30	50	86	62	20	18	24	60	66	18
			60	60	90	98	92	76	30	32	66	42	18	12	26	38	40	22
			30	32	52	70	62	50	6	24	48	18	8	6	10	12	8	8
		0-1	90	54	68	92	76	72	30	44	78	62	24	4	12	16	6	24
			60	16	26	30	20	10	12	24	46	38	22	8	14	38	26	18
			30	12	18	24	22	10	10	6	26	16	4	2	10	8	8	6
Aleatório	1.000	1	90	100	100	100	100	100	76	100	100	96	30	26	40	90	12	10
			60	100	100	100	100	98	44	94	96	88	18	16	22	74	20	6
			30	76	98	100	94	70	20	68	66	20	2	20	20	52	10	8
		0-1	90	94	100	100	100	92	48	100	100	74	14	10	18	64	24	6
			60	62	98	100	74	46	6	8	18	12	10	12	26	52	12	6
			30	38	84	98	66	22	22	60	40	20	4	2	12	8	14	6
	200	1	90	44	86	96	80	36	14	44	70	38	18	10	18	42	6	12
			60	56	88	100	88	40	8	28	62	28	4	10	12	30	12	12
			30	22	44	62	32	20	12	16	36	10	4	6	10	24	6	6
		0-1	90	48	84	100	72	38	10	12	32	8	0	20	26	40	16	2
			60	34	62	90	52	28	8	22	32	16	2	4	8	4	10	2
			30	8	16	18	8	8	4	8	4	0	6	2	14	12	0	2

M₁ a M₅ = marcas moleculares; grau de dominância 1 = dominância completa e 0-1 = dominância parcial; h² = herdabilidade em porcentagem; e * = localização do QTL.

Tabela 9 – Teste para dominância no loco do QTL em porcentagem de testes significativos a 5% no modelo de regressão quadrático, para cada marca molecular, utilizando o método da marca simples, em função da característica crescimento líquido relativo (%) com direção de dominância negativa, nível de saturação do mapa, tamanho do composto, grau de dominância, herdabilidade e diferença de frequência de mesma marca entre os genitores

Nível de saturação do mapa	Tamanho do composto	Grau de dominância	h ²	Diferença mínima de frequência das marcas														
				1,0					0,9					0,8				
				M1	M2	M3 *	M4	M5	M1	M2	M3 *	M4	M5	M1	M2	M3 *	M4	M5
Alto	1.000	1	90	100	100	100	100	96	100	100	100	100	76	62	94	80	74	
			60	100	100	100	100	84	100	100	100	100	50	48	82	62	60	
			30	100	100	100	100	62	84	100	96	88	32	20	48	40	36	
		0-1	90	100	100	100	100	66	92	100	100	100	26	32	50	34	28	
			60	98	98	100	100	66	88	100	100	88	8	8	28	28	24	
			30	96	100	100	100	60	88	100	100	88	22	12	38	32	22	
	200	1	90	98	100	100	100	40	64	86	72	22	14	22	38	32	26	
			60	92	98	100	98	26	42	64	54	12	14	8	26	26	18	
			30	66	74	92	88	14	30	46	28	6	6	6	10	12	8	
		0-1	90	96	100	100	100	24	46	76	44	12	10	18	18	8	18	
			60	86	92	100	100	10	4	6	2	2	10	12	4	12	18	
			30	36	34	44	44	6	14	20	22	8	6	4	4	2	6	
Aleatório	1.000	1	90	100	100	100	100	32	88	96	78	22	28	60	100	40	14	
			60	100	100	100	100	64	100	100	96	38	18	44	86	20	12	
			30	92	100	100	100	12	38	52	20	8	14	18	62	18	2	
		0-1	90	98	100	100	100	24	50	66	48	26	18	28	92	24	10	
			60	84	100	100	100	40	84	94	72	16	20	26	72	18	6	
			30	96	100	100	100	12	22	28	10	10	6	16	44	12	12	
	200	1	90	32	44	66	34	22	28	56	86	36	2	16	22	46	14	12
			60	70	96	100	90	62	24	46	74	30	14	16	28	12	8	
			30	52	82	80	52	30	12	34	32	8	2	8	12	12	10	6
		0-1	90	44	76	94	70	36	24	54	78	28	4	10	24	36	10	6
			60	12	28	42	22	20	14	32	54	12	2	10	8	10	4	6
			30	12	14	12	10	10	4	18	32	8	8	12	12	18	2	4

M₁ a M₅ = marcas moleculares; grau de dominância 1 = dominância completa e 0-1 = dominância parcial; h² = herdabilidade em porcentagem; e * = localização do QTL.

4.2.3- Avaliação dos efeitos de substituição de marca e dos desvios de dominância no loco do marcador

As análises de regressão nos modelos linear e quadrático possibilitaram a estimação dos efeitos de substituição de marca (α_M) e dos desvios de dominância (d_M) nos locos marcadores do QTL, respectivamente.

Em relação ao efeito de substituição da marca, quando se tem interesse em aumentar o valor fenotípico do caráter a seleção deve ser feita a favor das marcas com sinal positivo, e quando o interesse for e diminuir o valor fenotípico a seleção deve ser feita a favor das marcas com sinal negativo. Em virtude do interesse no aumento da expressão para os três caracteres idealizados, o melhorista deve praticar seleção contra os marcadores na região do QTL com sinal negativo para este efeito (Tabelas 10, 11 e 12).

Na avaliação da magnitude dos efeitos de substituição da marca, para todos os marcadores na região do QTL, foram verificados efeitos de magnitudes semelhantes entre as marcas para cada situação e tendência de maior efeito para as marcas localizadas mais próximas ao QTL, com raras exceções. Portanto, pode-se afirmar que a magnitude do efeito de substituição de marca é determinada apenas pela distância entre os marcadores e o QTL, uma vez que foram estas que apresentaram maior efeito (Tabelas 10, 11 e 12).

Quanto à avaliação dos efeitos dos desvios de dominância dos marcadores na região do QTL, para a característica produção determinada por QTL parcialmente positivo, verificou-se que os efeitos dos desvios de dominância estimados nos locos marcadores foram consistentes em revelar a direção de dominância no loco do QTL em todos os compostos (Tabela 10). Resultado equivalente ocorreu com QTL favorável dominante positivo apenas em populações de mapeamento equivalente à F_2 , com 1.000 genotipados. Para a característica capacidade de expansão, os sinais dos efeitos dos desvios de dominância variam entre positivo e negativo, para os locos marcadores na região do QTL, em cada situação (Tabela 11). Esse fato está relacionado à condição de direção de dominância bidirecional, estabelecida para esta característica. Assim, não foi possível determinar quais marcas têm efeito positivo ou negativo sobre a característica. Em relação à característica crescimento líquido relativo, determinada por QTL dominante recessivo, foi

observada, quase sem exceção, a adequação das estimativas dos efeitos de dominância nos locos dos marcadores em revelar a direção de dominância no loco do QTL, independentemente do grau de dominância (Tabela 12).

A obtenção de mapas moleculares e análise de QTL, a partir de populações endogâmicas ou de populações exogâmicas estruturadas em famílias, com base em análise de ligação clássica é bastante explorada. Porém, a obtenção de mapas genéticos e análise de QTL, a partir de populações não-endogâmicas e não-estruturadas em famílias, com base em desequilíbrio de fase gamética ainda não havia sido alvo específico de estudo. Assim, espera-se que este trabalho contribua para o avanço nos estudos de mapeamento genético e análise de QTL, em populações não-endogâmicas e não-estruturadas em famílias, com base em desequilíbrio de fase gamética.

Tabela 10 – Médias das estimativas dos efeitos de substituição da marca (α_M) no modelo de regressão linear e dos desvios de dominância no loco do marcador (d_M), modelo quadrático, utilizando o método da marca simples, em função da característica produção (g/planta) com direção de dominância positiva, nível de saturação do mapa, tamanho do composto, grau de dominância, herdabilidade e diferença de frequência de mesma marca entre os genitores

Nível de saturação do mapa	Tamanho do composto	Grau de dominância	Efeitos	h^2	Diferença mínima de frequência das marcas														
					1,0					0,9					0,8				
					M1	M2	M3 *	M4	M5	M1	M2	M3 *	M4	M5	M1	M2	M3 *	M4	M5
Alto	1.000	1	α_M	90	-3,81	-4,24	-4,64	-4,53	4,19	-4,34	5,15	6,08	6,02	5,17	5,55	-5,23	6,14	5,96	6,96
				60	-3,83	-4,25	-4,63	-4,54	4,17	-4,37	5,15	6,06	6,03	5,18	5,45	-5,14	6,00	5,94	6,98
				30	-3,65	-4,07	-4,48	-4,38	4,08	-4,37	5,05	5,95	5,91	5,11	5,43	-5,02	6,03	5,62	6,63
			d_M	90	4,90	5,63	6,35	6,08	5,30	-1,24	-1,89	-2,82	-3,15	-2,47	-1,59	-1,39	-2,16	-2,03	-2,73
				60	4,83	5,54	6,26	6,02	5,25	-1,22	-1,78	-2,85	-3,12	-2,43	-1,66	-1,40	-2,18	-1,91	-2,70
				30	4,92	5,72	6,47	6,19	5,38	-1,36	-2,08	-3,01	-3,44	-2,61	-1,40	-1,53	-2,13	-2,17	-3,07
		0-1	α_M	90	-4,29	-4,76	-5,21	-5,10	4,71	-3,69	4,45	5,49	5,64	4,72	3,98	-3,54	4,28	4,18	5,59
				60	-4,27	-4,76	-5,23	-5,09	4,69	-3,54	4,28	5,26	5,43	4,55	4,70	-4,30	5,15	4,99	6,26
				30	-4,27	-4,73	-5,14	-5,06	4,66	-3,75	4,49	5,62	5,75	4,76	3,19	-2,80	3,40	3,34	4,93
			d_M	90	2,14	2,47	2,80	2,73	2,41	0,83	0,97	1,40	1,44	1,08	1,34	1,05	1,41	1,39	1,97
				60	1,83	2,06	2,29	2,32	2,05	1,03	1,40	2,08	2,25	1,68	0,40	0,15	0,40	0,31	0,59
				30	2,99	3,39	3,88	3,66	3,11	0,78	1,17	1,88	1,83	1,36	1,80	1,69	2,30	2,40	3,18
	200	1	α_M	90	8,74	9,46	10,12	9,84	9,23	-1,51	2,13	-3,26	-2,96	-1,69	3,35	3,74	5,36	5,66	3,74
				60	8,77	9,50	10,06	9,82	9,30	-1,48	2,05	-3,23	-2,97	-1,68	3,25	3,71	5,23	5,63	3,90
				30	8,90	9,56	10,30	10,13	9,63	-1,65	2,25	-3,47	-3,38	-1,62	3,45	3,76	5,68	6,31	3,90
			d_M	90	-4,46	-5,00	-5,91	-5,56	-4,99	-1,34	-1,73	-2,76	-2,27	-0,92	-0,83	-1,18	-1,80	-2,14	-1,00
				60	-4,37	-5,03	-5,87	-5,61	-4,98	-1,42	-1,99	-2,79	-2,18	-0,84	-0,85	-1,05	-1,81	-2,13	-0,61
				30	-4,50	-5,03	-6,04	-5,48	-4,80	-1,00	-1,62	-2,32	-2,16	-0,46	-0,70	-1,23	-1,41	-1,35	-1,54
		0-1	α_M	90	9,04	9,73	10,49	10,15	9,54	-2,50	3,19	-4,40	-3,90	-2,66	2,04	2,26	3,47	3,92	2,40
				60	8,89	9,57	10,31	10,04	9,39	-2,97	3,69	-4,91	-4,27	-3,03	2,04	2,27	3,42	3,86	2,28
				30	8,70	9,61	10,33	10,18	9,44	-3,04	3,85	-4,75	-4,15	-3,21	2,55	2,97	4,55	4,80	3,30
			d_M	90	3,02	3,73	4,15	3,81	3,18	0,68	0,60	1,01	0,93	0,60	0,90	0,90	1,88	2,05	0,92
				60	1,72	2,28	2,52	2,24	1,77	1,49	1,68	2,62	2,46	1,76	1,13	0,98	1,81	2,05	1,07
				30	2,51	3,10	3,71	3,61	3,05	2,68	2,66	4,46	3,64	2,10	0,60	0,25	1,28	1,16	1,04

Continua ...

Tabela 10 - Continuação

Aleatório	1.000	1	α_M	90	-6,16	-7,53	8,34	7,31	6,08	-4,50	5,73	5,89	-5,69	4,10	1,61	1,93	2,61	1,28	0,62
				60	-6,14	-7,50	8,30	7,25	6,01	-4,46	5,72	5,91	-5,80	4,17	1,58	1,93	2,65	1,26	0,61
				30	-6,27	-7,68	8,38	7,48	6,18	-4,62	5,84	6,02	-5,71	4,25	1,60	1,91	2,65	1,35	0,71
		d_M	90	4,70	6,87	8,17	6,03	4,41	-1,54	-2,64	-2,77	-2,34	-1,23	0,51	0,88	1,45	0,37	0,15	
			60	4,74	6,98	8,21	5,87	4,33	-1,48	-2,58	-2,66	-2,33	-1,34	0,54	1,04	1,47	0,38	-0,01	
			30	4,71	7,07	8,31	6,04	4,39	-1,61	-2,70	-2,90	-2,56	-1,35	0,57	0,92	1,47	0,59	0,55	
	0-1	α_M	90	-6,56	-7,99	8,84	7,77	6,46	-2,98	3,98	3,85	-4,25	2,84	1,95	2,43	3,27	1,62	0,89	
			60	-6,55	-7,93	8,77	7,71	6,41	-3,34	4,38	4,28	-4,58	3,13	2,17	2,77	3,57	1,67	1,00	
			30	-6,61	-8,04	8,88	7,86	6,62	-2,87	4,06	3,95	-4,35	2,98	2,04	2,65	3,56	1,74	0,91	
		d_M	90	3,11	4,76	5,71	4,20	3,04	1,25	2,17	2,25	2,28	0,90	0,47	0,71	1,00	0,40	0,34	
			60	3,91	5,96	7,27	5,25	3,86	0,81	1,33	1,15	1,21	0,50	0,20	0,45	0,48	0,25	0,33	
			30	1,46	2,00	2,38	2,02	1,74	0,85	1,90	2,09	2,18	0,96	0,38	0,49	0,49	0,49	0,32	
	200	1	α_M	90	7,24	-8,81	9,66	8,54	7,22	4,32	6,61	7,59	-5,73	4,12	1,84	-2,34	3,49	2,70	1,52
				60	7,47	-9,02	9,83	8,38	7,03	4,19	6,35	7,39	-5,64	3,98	1,81	-2,37	3,48	2,63	1,57
				30	7,53	-9,04	9,94	8,87	7,43	4,24	6,44	7,32	-5,44	3,99	1,75	-2,25	3,42	2,65	1,48
		d_M	90	-3,97	-5,85	-6,81	-4,86	-3,21	-0,88	-1,06	-1,27	-0,68	-0,33	-0,69	-1,17	-1,81	-0,98	-0,53	
			60	-4,22	-6,14	-7,08	-5,05	-3,09	-1,01	-1,40	-1,57	-0,97	-0,88	-0,60	-1,27	-2,07	-1,21	-0,78	
			30	-3,85	-5,96	-6,93	-4,87	-3,31	-1,09	-1,18	-1,52	-0,49	-0,12	-0,56	-0,97	-1,71	-1,21	-0,80	
0-1	α_M	90	7,77	-9,38	10,22	8,82	7,29	3,26	5,20	6,08	-4,40	3,18	0,64	-0,95	1,86	1,59	0,67		
		60	7,66	-9,27	10,07	8,71	7,41	3,11	5,00	5,84	-4,13	3,09	0,89	-1,17	2,19	1,69	0,75		
		30	7,82	-9,42	10,32	8,79	7,29	0,83	1,34	1,58	-1,20	0,90	0,85	-1,19	1,94	1,62	0,88		
d_M	90	3,81	5,55	6,55	4,83	3,27	0,52	1,38	1,85	1,40	0,81	0,53	0,65	1,64	0,99	0,15			
	60	4,19	6,03	6,97	5,04	3,32	0,92	1,87	2,70	1,59	1,10	0,29	0,37	1,15	0,70	0,21			
	30	2,90	4,05	4,73	3,59	2,58	0,26	0,77	1,05	0,62	0,31	0,34	0,33	1,29	0,72	0,30			

M₁ a M₅ = marcas moleculares; grau de dominância 1 = dominância completa e 0-1 = dominância parcial; h² = herdabilidade em porcentagem; e * = localização do QTL.

Tabela 11 – Médias das estimativas dos efeitos de substituição da marca (α_M) no modelo de regressão linear e dos desvios de dominância no loco do marcador (d_M), modelo quadrático, utilizando o método da marca simples, em função da característica capacidade de expansão (mg/l) com direção de dominância bidirecional, nível de saturação do mapa, tamanho do composto, grau de dominância, herdabilidade e diferença de frequência de mesma marca entre os genitores

Nível de saturação do mapa	Tamanho do composto	Grau de dominância	Efeitos	h^2	Diferença mínima de frequência das marcas														
					1,0					0,9					0,8				
					M1	M2	M3 *	M4	M5	M1	M2	M3 *	M4	M5	M1	M2	M3 *	M4	M5
Alto	1.000	1	α_M	90	-1,21	-1,35	-1,48	-1,44	1,33	-1,46	1,70	1,99	1,94	1,67	1,78	-1,67	1,98	1,90	2,21
				60	-1,21	-1,35	-1,48	-1,44	1,34	-1,45	1,69	1,97	1,94	1,68	0,95	-0,82	0,96	0,92	1,36
				30	-1,23	-1,38	-1,50	-1,47	1,37	-0,81	0,95	1,20	1,27	1,01	1,79	-1,65	1,99	1,88	2,18
			d_M	90	-1,12	-1,40	-1,69	-1,62	-1,42	-0,43	-0,74	-1,19	-1,35	-1,04	-0,53	-0,46	-0,66	-0,69	-0,98
				60	-1,17	-1,46	-1,72	-1,64	-1,43	-0,52	-0,82	-1,27	-1,38	-1,03	0,40	0,31	0,50	0,59	0,77
				30	-0,86	-1,11	-1,37	-1,29	-1,08	0,34	0,59	1,04	1,23	0,89	-0,27	-0,34	-0,56	-0,54	-0,79
		0-1	α_M	90	-1,21	-1,35	-1,48	-1,44	1,33	-1,35	1,61	1,95	1,96	1,66	1,65	-1,53	1,82	1,77	2,15
				60	-1,20	-1,34	-1,47	-1,45	1,33	-1,41	1,69	2,03	2,03	1,72	1,04	-0,91	1,11	1,11	1,63
				30	-1,24	-1,38	-1,50	-1,48	1,37	-1,40	1,68	1,98	1,96	1,70	1,66	-1,53	1,84	1,80	2,26
			d_M	90	0,81	0,98	1,14	1,08	0,90	0,00	-0,13	-0,26	-0,39	-0,34	-0,10	-0,10	-0,17	-0,16	-0,21
				60	-0,05	-0,00	0,03	0,06	0,04	-0,18	-0,37	-0,61	-0,70	-0,56	0,46	0,40	0,65	0,72	0,93
				30	0,53	0,64	0,77	0,74	0,60	-0,19	-0,37	-0,61	-0,73	-0,49	-0,06	-0,06	-0,22	-0,24	-0,08
	200	1	α_M	90	2,61	2,83	3,05	2,96	2,80	-0,88	1,09	-1,43	-1,20	-0,84	0,48	0,55	0,90	1,07	0,70
				60	2,64	2,84	3,10	2,99	2,78	-0,85	1,07	-1,40	-1,24	-0,89	0,48	0,56	0,88	1,07	0,73
				30	2,53	2,74	2,98	2,93	2,75	-0,39	0,60	-1,05	-0,97	-0,58	1,19	1,38	2,04	2,43	1,56
			d_M	90	-1,60	-1,81	-2,13	-1,98	-1,73	0,64	0,77	1,24	1,03	0,57	0,26	0,38	0,83	0,98	0,33
				60	1,20	1,65	2,06	1,87	1,48	0,79	0,83	1,27	1,11	0,63	0,36	0,44	0,83	1,01	0,59
				30	-1,39	-1,73	-2,15	-1,98	-1,78	-0,57	-0,87	-1,53	-1,15	-0,58	-0,22	-0,47	-0,63	-0,67	-0,36
		0-1	α_M	90	2,90	3,13	3,37	3,27	3,08	-0,86	1,09	-1,43	-1,24	-0,85	1,03	1,16	1,68	1,79	1,12
				60	2,96	3,18	3,41	3,29	3,08	-0,75	0,98	-1,33	-1,10	-0,73	0,66	0,72	1,16	1,28	0,77
				30	2,91	3,12	3,36	3,27	3,06	-0,73	0,91	-1,27	-1,16	-0,76	0,94	1,00	1,47	1,58	1,00
			d_M	90	-0,84	-0,91	-1,10	-1,05	-0,94	0,59	0,71	1,14	0,98	0,61	-0,04	-0,16	-0,21	-0,21	-0,14
				60	0,46	0,62	0,72	0,63	0,40	0,41	0,53	0,97	0,90	0,68	0,29	0,31	0,66	0,69	0,20
				30	0,75	1,05	1,10	0,98	0,92	0,47	0,46	0,86	0,70	0,36	0,04	0,09	0,30	0,23	0,28

Continua ...

Tabela 11 - Continuação

Aleatório	1.000	1	α_M	90	-1,99	-2,42	2,67	2,35	1,98	-0,96	1,28	1,24	-1,38	0,95	0,94	1,14	1,45	0,75	0,46
				60	-1,98	-2,40	2,66	2,34	1,97	-0,93	1,24	1,22	-1,35	0,92	0,90	1,13	1,45	0,79	0,50
				30	-1,99	-2,43	2,66	2,32	1,96	-1,32	1,67	1,71	-1,65	1,20	0,52	0,61	0,80	0,46	0,31
		d_M	90	-1,02	-1,64	-1,99	-1,42	-1,01	0,39	0,75	0,76	0,68	0,21	-0,21	-0,28	-0,54	-0,12	0,03	
			60	-1,09	-1,67	-2,09	-1,54	-1,02	0,35	0,73	0,73	0,74	0,22	-0,14	-0,27	-0,56	-0,26	-0,19	
			30	-1,02	-1,47	-1,91	-1,37	-0,90	-0,36	-0,72	-0,72	-0,38	-0,10	0,26	0,35	0,61	0,17	0,15	
	0-1	α_M	90	-2,17	-2,64	2,92	2,56	2,15	-1,39	1,79	1,83	-1,77	1,26	0,87	1,10	1,42	0,80	0,51	
			60	-2,14	-2,62	2,89	2,53	2,11	-1,25	1,61	1,64	-1,65	1,18	0,60	0,75	1,01	0,51	0,30	
			30	-2,10	-2,62	2,89	2,53	2,14	-1,08	1,43	1,40	-1,45	0,98	0,80	0,97	1,26	0,64	0,36	
		d_M	90	-0,78	-1,16	-1,39	-1,01	-0,64	-0,34	-0,65	-0,69	-0,44	-0,19	-0,09	-0,15	-0,36	-0,19	-0,12	
			60	0,59	0,90	1,07	0,67	0,44	-0,05	-0,15	-0,17	-0,11	-0,07	0,09	0,23	0,40	0,12	0,04	
			30	0,65	1,06	1,30	0,83	0,54	0,30	0,59	0,52	0,30	0,04	-0,14	-0,15	-0,20	0,04	-0,02	
	200	1	α_M	90	2,34	-2,86	3,11	2,74	2,32	4,41	6,13	6,86	-5,08	3,57	0,59	-0,75	1,12	0,88	0,54
				60	2,33	-2,85	3,12	2,70	2,33	0,90	1,40	1,65	-1,11	0,85	0,57	-0,73	1,11	0,90	0,50
				30	2,27	-2,81	3,12	2,61	2,18	1,59	2,33	2,64	-1,88	1,40	0,17	-0,17	0,44	0,39	0,07
		d_M	90	-0,85	-1,45	-1,78	-1,23	-0,75	0,22	0,42	0,60	0,57	-0,15	-0,14	-0,29	-0,53	-0,29	-0,14	
			60	-1,31	-2,04	-2,48	-1,79	-1,13	0,25	0,81	1,26	0,73	0,28	-0,16	-0,31	-0,62	-0,29	-0,23	
			30	-1,35	-1,77	-2,20	-1,55	-1,05	-0,61	-1,06	-1,41	-0,85	-0,20	0,24	0,31	0,72	0,41	-0,10	
0-1	α_M	90	2,47	-2,99	3,27	2,86	2,40	1,34	2,06	2,39	-1,78	1,31	0,27	-0,37	0,69	0,59	0,29		
		60	2,42	-2,93	3,24	2,77	2,28	1,15	1,77	2,05	-1,47	1,01	0,37	-0,45	0,81	0,66	0,40		
		30	2,46	-2,93	3,23	2,78	2,23	1,06	1,74	2,03	-1,59	1,15	0,37	-0,45	0,81	0,66	0,40		
d_M	90	-0,87	-1,37	-1,64	-1,15	-0,73	-0,24	-0,37	-0,49	-0,27	-0,11	0,14	0,15	0,41	0,16	-0,10			
	60	0,81	1,36	1,62	1,24	0,87	0,04	0,41	0,58	0,37	0,15	-0,02	-0,07	0,07	0,06	0,02			
	30	0,49	0,91	0,99	0,59	0,44	-0,01	-0,02	0,00	0,09	0,07	-0,02	-0,07	0,07	0,06	0,02			

M_1 a M_5 = marcas moleculares; grau de dominância 1 = dominância completa e 0-1 = dominância parcial; h^2 = herdabilidade em porcentagem; e * = localização do QTL.

Tabela 12 – Médias das estimativas dos efeitos de substituição da marca (α_M) no modelo de regressão linear e dos desvios de dominância no loco do marcador (d_M), modelo quadrático, utilizando o método da marca simples, em função da característica crescimento líquido relativo (%) com direção de dominância negativa, nível de saturação do mapa, tamanho do composto, grau de dominância, herdabilidade e diferença de frequência de mesma marca entre os genitores

Nível de saturação do mapa	Tamanho do composto	Grau de dominância	Efeitos	h^2	Diferença mínima de frequência das marcas														
					1,0					0,9					0,8				
					M1	M2	M3 *	M4	M5	M1	M2	M3 *	M4	M5	M1	M2	M3 *	M4	M5
Alto	1.000	1	α_M	90	-3,47	-3,85	-4,22	-4,12	3,79	-4,33	5,05	5,91	5,78	5,00	5,32	-5,07	5,94	5,69	6,68
				60	-3,56	-3,95	-4,30	-4,18	3,84	-4,31	5,06	5,92	5,76	5,00	5,34	-5,10	5,91	5,72	6,71
				30	-3,47	-3,88	-4,29	-4,18	3,82	-4,29	5,05	5,96	5,86	5,06	5,32	-5,06	5,89	5,71	6,88
			d_M	90	-4,52	-5,22	-5,96	-5,66	-4,99	-1,97	-2,78	-4,04	-4,36	-3,37	-1,73	-1,51	-2,30	-2,10	-3,07
				60	-4,62	-5,31	-5,99	-5,73	-5,06	-1,94	-2,79	-4,05	-4,34	-3,46	-1,70	-1,47	-2,36	-2,18	-3,08
				30	-4,54	-5,12	-5,85	-5,61	-4,92	-2,02	-2,60	-3,85	-4,26	-3,40	-1,90	-1,47	-2,36	-2,13	-3,39
		0-1	α_M	90	-3,48	-3,87	-4,25	-4,14	3,82	-4,13	4,89	5,85	5,86	5,04	4,98	-4,67	5,55	5,38	6,56
				60	-3,49	-3,86	-4,23	-4,12	3,81	-4,29	5,08	6,05	6,04	5,18	5,04	-4,77	5,54	5,43	6,59
				30	-3,47	-3,92	-4,32	-4,22	3,87	-4,57	5,33	6,30	6,20	5,31	5,42	-5,25	6,20	5,94	7,15
			d_M	90	-3,21	-3,82	-4,43	-4,20	-3,67	-0,93	-1,37	-1,98	-2,14	-1,67	-0,84	-0,73	-1,04	-0,87	-1,29
				60	-1,76	-1,95	-2,12	-2,00	-1,78	-1,25	-1,84	-2,57	-2,80	-2,08	-0,50	-0,55	-1,06	-1,01	-1,54
				30	-2,53	-3,03	-3,59	-3,48	-3,12	-1,85	-2,76	-3,95	-4,33	-3,15	-1,39	-1,32	-1,92	-1,87	-2,23
	200	1	α_M	90	7,64	8,30	8,91	8,68	8,17	-0,81	1,29	-2,39	-2,22	-0,99	3,41	3,87	5,44	5,58	3,78
				60	7,73	8,33	8,93	8,61	8,06	-0,66	1,31	-2,38	-2,23	-0,88	3,44	3,94	5,46	5,61	3,88
				30	7,54	8,18	8,76	8,51	7,96	-0,74	1,29	-2,25	-2,11	-0,80	3,31	3,72	5,28	5,76	3,94
			d_M	90	-6,01	-6,70	-7,77	-7,33	-6,49	-2,24	-2,91	-4,42	-3,71	-1,74	-1,20	-1,45	-2,31	-2,72	-1,29
				60	-6,08	-6,64	-7,75	-7,34	-6,41	-1,94	-2,61	-4,18	-3,67	-1,67	-1,36	-1,41	-2,45	-2,92	-1,00
				30	-5,81	-6,60	-7,85	-7,45	-6,51	-2,06	-3,10	-4,68	-4,18	-1,93	-1,50	-1,79	-2,56	-2,47	-1,04
		0-1	α_M	90	7,65	8,30	8,88	8,64	8,12	-1,43	1,96	-3,07	-2,84	-1,53	3,01	3,37	4,86	5,18	3,39
				60	7,43	8,04	8,62	8,42	7,89	-1,78	2,37	-3,30	-2,95	-1,71	3,14	3,47	5,04	5,32	3,41
				30	7,81	8,26	8,98	8,65	7,97	-1,28	1,85	-3,02	-2,83	-1,74	2,60	2,94	4,40	4,91	2,93
			d_M	90	-4,96	-5,64	-6,68	-6,29	-5,60	-1,44	-1,96	-2,98	-2,39	-1,02	-0,64	-0,92	-1,09	-1,24	-0,49
				60	-4,89	-5,64	-6,55	-6,28	-5,53	-0,43	-0,72	-0,99	-1,06	-0,34	-0,53	-0,55	-0,65	-0,98	-0,73
				30	-3,41	-3,51	-4,20	-3,89	-3,36	-1,30	-1,43	-2,91	-2,66	-0,63	-0,30	-0,42	0,08	-0,10	-0,45

Continua ...

Tabela 12 – Continuação

Aleatório	1.000	1	α_M	90	-5,76	-6,96	7,66	6,79	5,74	-2,99	3,94	3,88	-4,25	2,93	3,04	3,67	4,64	2,72	1,81
				60	-5,67	-6,94	7,68	6,79	5,71	-2,71	3,59	3,47	-3,88	2,58	3,07	3,68	4,62	2,73	1,82
				30	-5,47	-6,60	7,29	6,41	5,47	-3,23	4,12	3,97	-4,26	2,75	3,03	3,81	4,74	2,74	1,80
			d_M	90	-4,45	-6,36	-7,49	-5,61	-4,05	0,64	1,52	1,69	1,26	0,30	-0,82	-1,10	-2,26	-1,02	-0,65
				60	-4,42	-6,26	-7,37	-5,47	-3,95	1,38	2,72	2,71	2,36	1,06	-0,77	-1,16	-2,27	-1,10	-0,58
				30	-4,11	-6,01	-7,09	-5,44	-3,95	0,63	1,61	1,73	1,23	0,44	-0,91	-0,97	-2,38	-1,16	-0,44
		0-1	α_M	90	-6,25	-7,62	8,39	7,40	6,21	-4,19	5,42	5,50	-5,45	3,87	2,74	3,34	4,24	2,36	1,49
				60	-6,13	-7,52	8,29	7,35	6,15	-3,30	4,39	4,33	-4,80	3,19	2,68	3,32	4,18	2,38	1,53
				30	-6,14	-7,46	8,23	7,32	6,05	-4,26	5,55	5,57	-5,31	3,65	2,79	3,33	4,19	2,44	1,60
			d_M	90	-2,61	-3,87	-4,59	-3,51	-2,49	-0,50	-0,90	-1,02	-0,97	-0,63	-0,51	-0,71	-1,57	-0,69	-0,44
				60	-2,09	-2,95	-3,50	-2,71	-1,91	0,92	1,99	2,15	1,73	0,40	-0,46	-0,75	-1,63	-0,73	-0,48
				30	-4,10	-5,85	-6,87	-5,22	-3,62	-0,69	-1,30	-1,30	-0,91	-0,60	-0,36	-0,70	-1,75	-0,62	-0,41
	200	1	α_M	90	7,64	-9,25	10,17	8,83	7,31	4,55	6,66	7,58	-5,90	4,15	1,96	-2,44	3,57	2,72	1,53
				60	7,20	-8,95	9,93	8,72	7,41	4,44	6,64	7,56	-5,82	4,02	2,06	-2,40	3,58	2,76	1,63
				30	7,14	-8,66	9,44	8,18	7,01	4,52	6,31	7,22	-5,74	4,10	1,88	-2,45	3,79	2,41	1,35
			d_M	90	1,69	2,37	2,82	2,15	1,58	-1,89	-3,50	-4,45	-2,76	-1,55	-0,82	-1,24	-2,18	-1,60	-0,88
				60	-5,22	-7,61	-8,86	-6,57	-4,45	-1,89	-3,50	-4,62	-2,87	-1,77	-0,78	-1,34	-2,18	-1,38	-0,59
				30	-5,10	-7,19	-7,88	-5,81	-3,72	-2,02	-3,44	-4,29	-1,90	-1,24	-0,63	-1,26	-1,93	-2,02	-1,58
		0-1	α_M	90	7,42	-9,02	9,90	8,67	7,31	4,18	6,20	7,11	-5,45	3,82	1,65	-2,18	3,29	2,48	1,35
				60	7,21	-8,85	9,81	8,37	6,95	4,10	6,07	7,02	-5,24	3,86	1,33	-1,67	2,81	2,23	1,35
				30	7,44	-8,94	9,79	8,63	7,46	4,37	6,25	7,13	-5,13	3,72	1,83	-2,26	3,33	2,66	1,38
			d_M	90	-2,38	-3,60	-4,22	-3,12	-1,93	-1,54	-2,81	-3,57	-2,13	-1,24	-0,57	-0,94	-1,50	-0,92	-0,48
				60	-1,73	-2,47	-2,79	-2,15	-1,46	-1,27	-2,59	-3,34	-1,67	-0,85	-0,33	-0,53	-0,74	-0,66	-0,48
				30	-1,28	-1,68	-1,76	-1,61	-1,27	-1,57	-2,78	-3,79	-2,02	-1,42	-0,38	-0,85	-1,41	-0,41	-0,15

M₁ a M₅ = marcas moleculares; grau de dominância 1 = dominância completa e 0-1 = dominância parcial; h² = herdabilidade em porcentagem; e* = localização do QTL.

5- CONCLUSÕES

A população mais eficiente para o mapeamento é um composto de segunda geração equivalente a uma população de mapeamento F_2 . Porém, se estimar a ordem das marcas for tão ou até mais importante que a estimação do tamanho do grupo de ligação, compostos obtidos de populações que apresentarem diferença mínima de frequência de mesma marca nos genitores de até 0,8 também são eficientes para o mapeamento.

A eficiência na recuperação dos genomas paramétricos foi semelhante entre os compostos de tamanhos diferentes, dentro de cada nível de saturação do mapa. Assim, a utilização de amostras de composto com 200 indivíduos seria suficiente para resgatar as informações dos genomas paramétricos de forma satisfatória, sendo desnecessário avaliar e genotipar maior quantidade de indivíduos.

Considerando todas as diferenças mínimas de frequência de mesma marca entre os genitores, constatou-se a recuperação do genoma paramétrico quando foi utilizado nível aleatório de saturação do genoma.

O mapeamento por marca simples foi eficiente na detecção de QTL.

A contribuição de cada QTL para a variação fenotípica das características foi dependente da herdabilidade da característica, sendo da ordem de um terço de cada herdabilidade.

6- REFERÊNCIAS BIBLIOGRÁFICAS

- ALZATE-MARIN, A. L.; COSTA, M. R.; MENARIM H.; MOREIRA, M. A.; BARROS, E. G. Herança da resistência à Antracnose na cultivar de feijoeiro comum Cornell 49-242. **Fitopatologia Brasileira**, v. 28, n.3, p. 302-306, mai./jun. 2003.
- ARRIEL, N. H. C.; DI MAURO, A. O.; DI MAURO, S. M. Z.; BAKKE, O. A.; UNÊDA-TREVISOLI, S. H.; COSTA, M. M.; CAPELOTO, A.; CORRADO, A. R. Técnicas multivariadas na determinação da diversidade genética em gergelim usando marcadores RAPD. **Pesquisa Agropecuária Brasileira**, Brasília, v. 41, n. 5, p. 801-809, maio, 2006.
- AUSTIN, D. F.; LEE, M. Detection of quantitative trait loci for grain yield and yield components in maize across generations in stress and nonstress environments. **Crop Science**, v.38, p.1296-1308, 1998.
- BASTIANEL, M.; OLIVEIRA, A. C.; CRISTOFANI, M.; MACHADO, M. A. Diversidade genética entre híbridos de laranja-doce e tangor 'Murcott' avaliada por fAFLP e RAPD. **Pesquisa Agropecuária Brasileira**, Brasília, v. 41, n. 5, p. 779-784, maio, 2006.

- BORÉM, A. Aplicação dos marcadores moleculares no melhoramento de plantas. In: BORÉM, A., CAIXETA, E. T. (Ed.) **Marcadores moleculares**. Viçosa: UFV, 2006, p. 79-84.
- BORÉM, A.; MIRANDA, G. V. **Melhoramento de plantas**. 4ª ed. Viçosa: UFV, 2005, p.441-463.
- BRONDANI, C.; BRONDANI, R. P. V.; RANGEL, P. H. N.; FERREIRA, M. E. Development and mapping of *Oryza glumaepatula*-derived microsatellite markers in the interspecific cross *Oryza glumaepatula* x *O. sativa*. **Hereditas**, Lund, v. 134, n. 1, p. 59-71, maio, 2001.
- BRONDANI, C.; BRONDANI, R. P. V.; RANGEL, P. H. N. **Utilização de Marcadores Moleculares em Programas de Ampliação da Base Genética de Espécies Cultivadas**. Santo Antônio de Goiás : Embrapa Arroz e Feijão, 36 p. 2003.
- BURR, B.; BURR, F. A.; THOMPSON, K. H.; ALBERTSON, M.; STUBER, C. W. Gene mapping with recombinant inbreds in maize. **Genetics**, v. 118, p. 519-526, 1988.
- CAIXETA, E. T.; OLIVEIRA, A. C. B.; BRITO, G. G.; SAKIYAMA, N. S. Tipos de marcadores moleculares. In: BORÉM, A., CAIXETA, E. T. (Ed.) **Marcadores moleculares**. Viçosa: UFV, 2006. p. 9-78.
- CARDINAL, A. J.; LEE, M.; SHAROPORA, N.; WOODMAN-CLIKEMAN, W. L.; LONG, M. J. Genetic mapping and analysis of quantitative trait loci for resistance to stalk tunneling by the European corn bores in maize. **Crop Science**, v. 41, p. 835-845, 2001.
- CHING, A.; CALDWELL, K. S.; JUNG, M.; DOLAN, M., SMITH, O. S.; TINGEY, S.; MORGANTE, M.; RAFALSKI, A. J. SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. **BMC Genetics**, v. 3, p. 19, 2002.

- COELHO, A. S. G. Considerações gerais sobre a análise de QTL's. In: PINHEIRO, J. B. (ed.) **Análise de QTL no melhoramento de plantas: segunda jornada em Genética e melhoramento de plantas**. FUNAPE, Goiânia, 2000, p. 1-20.
- CRUZ, C. D. A informática no melhoramento genético. In: Nass, L. L.; Valois, A. C. C; Melo, I. S.; Valadares-Inglis, M. C. (Ed.) **Recursos genéticos & melhoramento**. Rondonópolis: Fundação MT, 2001. p. 1085-1118.
- CRUZ, C. D.; CARNEIRO, P. C. S. **Modelos biométricos aplicados ao melhoramento genético**. Vol. 2, Viçosa: Imprensa Universitária, 2003. 279p.
- ELANDT-JOHNSON, R. C. **Probability models and statistical methods in genetics**. New York, USA: John Wiley & Sons, Inc. 592p. 1971.
- FALEIRO, F. G.; RAGAGNIN, V. A.; SCHUSTER, I.; CORRÊA, R. X.; GOODGOD, P. I.; BROMMONSHENKEL, S. H.; MOREIRA, M. A.; BARROS, E. G. Mapeamento de genes de resistência do feijoeiro à ferrugem, antracnose e mancha-angular usando marcadores RAPD. **Fitopatologia Brasileira**, v. 28, n. 1, p. 59-66, jan./fev. 2003.
- FERREIRA, D. F. Uso de simulação no melhoramento. In: Nass, L. L.; Valois, A. C. C; Melo, I. S.; Valadares-Inglis, M. C. (Ed.) **Recursos genéticos & melhoramento**. Rondonópolis: Fundação MT, 2001. p. 1119-1142.
- FERREIRA, M. E., GRATTAPAGLIA, D. **Introdução ao uso de marcadores moleculares em análise genética**. Brasília, DF: EMBRAPA, CENARGEN, 1998. 220p.
- FLINT-GARCIA, S. A.; THORNSBERRY, J. M.; BUCKLER, E. S. Structure of linkage disequilibrium in plants. **Annual Review Plant of Biology**, v. 54, p. 357–374, 2003.

- GARCIA, A. F. **Mapeamento de QTL's para caracteres de importância agrônômica em duas populações F2 de milho tropical**. Campinas, SP: UNICAMP, 2003. 121p. Dissertação (Mestrado). Universidade Estadual de Campinas, Campinas.
- GLAZIER, A. M.; NADEAU, J. H.; AITMAN, T. J. Finding genes that underlie complex traits. **Science**, v. 298, p. 2345–2349, 2002.
- GORELICK, R.; LAUBICHLER, M. D. Decomposing multilocus linkage disequilibrium. **Genetics**, v. 166, p. 1581–1583, 2004.
- GUIMARAES, C. T.; HONEYCUTT, R. J.; SILLS, G. R.; SOBRAL, B. W. S. Genetic maps of *Saccharum officinarum* L. and *Saccharum robustum* Brandes & Jew. ex grassl. **Genetics and Molecular Biology**, v. 22, n. 1, 1999.
- GUIMARÃES, C. T., MOREIRA, M. A. Genética molecular aplicada ao melhoramento de plantas. In: Borém, A. (Ed.) **Melhoramento de espécies cultivadas**. 2ª ed. Viçosa: UFV, 2005. p. 865-895.
- GUPTA, P. K.; RUSTGI, S.; KULWAL, P. L. Linkage disequilibrium and association studies in higher plants: Present status and future prospects. **Plant Molecular Biology**, v. 57, p. 461-485, 2005.
- HALEY, C. S.; KNOTT, S. A. A simple regression method for mapping quantitative trait loci in the line cross using flanking markers. **Heredity**, v. 69, p. 315-324, 1992.
- JANSEN, R. C.; STAM, P. High resolution of quantitative traits into multiple loci via interval mapping. **Genetics**, v. 136, p. 1447-1455, 1994.
- JARDIM, S. N.; PAIVA, E.; MAGALHAES, J. V.; GUIMARAES, C. T.; PARENTONI, S. N.; SCHAFFERT, R. E. Mapping of aluminum QTLs in

maize. In: **50° Congresso Nacional de Genética**, 2004, Florianópolis - SC. Anais do 50° Congresso Nacional de Genética, 2004.

KAO, C. H.; ZENG, Z. B.; Teasdale, R. Multiple interval mapping for quantitative trait loci. **Genetics**, v. 152 p. 1203-1216, 1999.

KEMPTHORNE, O. **An introduction to genetic statistics**. New York: Wiley e Sons, 1957. 545p.

KRAAKMAN, A. T. W.; NIKS, R. E.; Van den BERG, P. M. M. M.; STAM, P.; Van EUWIJK, F.A. Linkage disequilibrium mapping of yield and yield stability in modern spring barley cultivars. **Genetics**, v. 168, p. 435-446, 2004.

KUMAR, S.; ECHT, C.; WILCOX, P. L.; RICHARDSON, T. E. Testing for linkage disequilibrium in the New Zealand radiata pine breeding population. **Theoretical and Applied Genetics**, v. 108, p. 292-298, 2004.

LANDER, E. S.; BOTSTEIN, D. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. **Genetics**, v. 121, p. 185-199, 1989.

LANZA, M. A.; GUIMARAES, C. T.; SCHUSTER, I. Aplicação de marcadores moleculares no melhoramento genético. **Informe Agropecuário**, Belo Horizonte, v. 21, n. 204, p. 97-108, maio/jun., 2000

LIMA, M. L. A. **Mapeamento de QTL para produção de grãos e caracteres de planta em milho tropical utilizando marcadores microssatélites**. Campinas, SP: UNICAMP, 2006. 140p. (Dissertação Doutorado). Universidade Estadual de Campinas, Campinas.

LITT, M., LUTY, J.A. A hypervariable microsatellite revealed by *in vitro* amplification of a dinucleotide repeat within the cardiac muscle actin gene. **American Journal Human Genetic**, v. 44, p. 398-401, 1989.

- LIU, B. H. **Statistical genomics: linkage, mapping, and QTL analysis**. Boca Raton, USA: CRC Press. 611p. 1998.
- LIU, K.; GOODMAN, M.; MUSE, S.; SMITH, J. S.; BUCKLER, E.; DOEBLEY, J. Genetic structure and diversity among maize inbred lines as inferred from DNA microsatellites. **Genetics**, v. 165, p. 2117–2128, 2003.
- LYNCH, M., WALSH, B. **Genetics and analysis of quantitative traits**. Sunderland, Massachusetts, USA: Sinauer Associates, Inc. 980p. 1998.
- MACHAY, T. F. C. The genetic architecture of quantitative traits. **Annual Review Genetics**, v.33, p. 303–39, 2001a.
- MACKAY, T. F. C. Quantitative trait loci in *Drosophila*. **Nature Review Genetics**, v. 2, p. 11-20, 2001b.
- MANGOLIN, C. A.; SOUZA JR, C. L.; GARCIA, A. A. F.; GARCIA, A. F.; SIBOV, S. T.; SOUZA, A. P. Mapping QTLs for kernel oil content in a tropical maize population. **Euphytica**, 137, p. 251–259, 2004.
- MATHIAS R. M.; SAGREDO, D. B.; KALAZICH, B. J. Uso de marcadores SSR para identificación de germoplasma de papa en el programa de mejoramiento de INIA de Chile. **Agricultura Técnica**, v. 67, n. 1, p. 3-15, mar. 2007.
- MIKLAS, P. N.; JOHNSON, W. C.; DELORME, R.; GEPTS, P. QTL conditioning physiological resistance and avoidance to white mold in dry bean. **Crop Science**, v. 41, p. 309-315, 2001.
- NJITI, V. N.; MEKSEM, K.; IQBAL, M. J.; JOHNSON, J. E.; KASSEM, M. A.; ZOBRIK, K. F.; KILO, V. Y.; LIGHFOOT, D. A. Common loci underlie field resistance to soybean sudden death syndrome in Forrest, Pyramid, Essex, and Douglas. **Theoretical and Applied Genetics**, v. 104, p. 294-300, 2002.

- NORDBORG, M.; TAVARÉ, S. Linkage disequilibrium: what history has to tell us? **Trends in Genetics**, v.18, p. 83–90, 2002.
- OLIVEIRA, R. P.; AGUILAR-VILDOSO, C. I.; CRISTOFANI, M.; MACHADO, M. A. Skewed RAPD markers in linkage maps of *Citrus*. **Genetics and Molecular Biology**, v. 27, n. 3, p. 437-441, 2004.
- PALAISSA, K. A.; MORGANTE, M.; WILLIAMS, M.; RAFALSKI, A. Contrasting effects of selection on sequence diversity and linkage disequilibrium at two phytoene synthase loci. **Plant Cell**, v. 15, p. 1795–1806, 2003.
- PEREIRA, M. G., PEREIRA, T. N. S. Marcadores moleculares no pré-melhoramento de plantas. In: BORÉM, A., CAIXETA, E. T. (Ed.) **Marcadores moleculares**. Viçosa: UFV, 2006. p. 85-106.
- RADMANN, E. B.; BIANCHI, V. J.; OLIVEIRA, R. P.; FACHINELLO, J. C. Caracterização e diversidade genética de cultivares de morangueiro. **Horticultura Brasileira**, v. 24, n. 1, jan./mar. 2006
- RAFALSKI, A. Applications of single nucleotide polymorphisms in crop genetics. **Current Opinion in Plant Biology**, v. 5, p. 94–100, 2002.
- SALLA, M. F. S.; RUAS, C. F.; RUAS, P. M.; CARPENTIERI-PÍPOLO, V. Uso de marcadores moleculares na análise da variabilidade genética em acerola (*Malpighia emarginata* D.C.). **Revista Brasileira de Fruticultura**, v.24, n.1, p.15-22, abr. 2002.
- SANTOS, M. A.; NICOLÁS, M. F.; HUNGRIA, M. Identificação de QTL associados à simbiose entre *Bradyrhizobium japonicum*, *B. elkanii* e soja. **Pesquisa Agropecuária Brasileira**, Brasília, v.41, n.1, p.67-75, jan. 2006.
- SAS System Release 9.00. **SAS Institute Inc.** Cary, NC, USA. 2002.

SCHUSTER, I.; CRUZ, C. D. **Estatística genômica aplicada a populações derivadas de cruzamentos controlados**. Viçosa: UFV. 568 p. 2004.

SIBOV, S. T.; SOUZA JR, C. L.; GARCIA, A. A. F.; GARCIA, A. F.; SILVA, A. R.; MANGOLIN, C. A. BENCHIMOL, L. L.; SOUZA, A. P. Molecular mapping in tropical maize (*Zea mays* L.) using microsatellite markers. 1. Map construction and localization of loci showing distorted segregation. **Hereditas**, p. 139, 96–106, 2003a.

SIBOV, S. T.; SOUZA JR, C. L.; GARCIA, A. A. F.; SILVA, A. R.; GARCIA, A. F.; MANGOLIN, C. A.; BENCHIMOL, L. L.; SOUZA, A. P. Molecular mapping in tropical maize (*Zea mays* L.) using microsatellite markers. 2. Quantitative trait loci (QTL) for grain yield and related traits. **Hereditas**, n. 139, p. 107–115, 2003b.

SILVA, L. C. **Simulação do tamanho da população e da saturação do genoma para mapeamento genético de RILs**. Viçosa, 2005. 120 p. Dissertação (Mestrado), Universidade Federal de Viçosa, Viçosa.

SOUZA, A. P. Biologia molecular aplicada ao melhoramento. In: Nass, L. L.; Valois, A. C. C; Melo, I. S.; Valadares-Inglis, M. C. (Ed.) **Recursos genéticos & melhoramento**. Rondonópolis: Fundação MT, 2001. p. 939-966.

SPIELMAN, R. S.; MCGINNIS, R. E.; EWENS, W. J. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). **American Journal of Human Genetics**, v. 52, p. 506–513, 1993.

STUBER, C. W.; EDWARDS, M. D.; WENDEL, J. F. Molecular marker-facilitated investigations of quantitative trait loci in maize. II. Factors influencing yield and its component traits. **Crop Science**, v. 27, p. 639-648, 1987.

VIANA, J. M. S. Quantitative genetics theory for non-inbred populations in linkage disequilibrium. **Genetics and Molecular Biology**, v. 27, nº 4, p. 594-601, 2004.

VIANA, J. M. S. Programa para análises de dados moleculares e quantitativos RealBreeding. Viçosa: UFV, 2007 (Software em desenvolvimento).

ZENG, Z. B. Precision mapping of quantitative trait loci. **Genetics**, v. 136, p. 1457-1468, 1994.

ZHU, Y. L.; SONG, Q. J.; HYTEN, D. L.; Van TASSELL, C. P.; MATUKUMALLI, L. K.; GRIMM, D. R.; HYATT, S. M.; FICKUS, E. W.; YOUNG, N. D.; CREGAN, P. B. Single nucleotide polymorphisms in soybean. **Genetics**, v. 163, p. 1123–1134, 2003.