

RODRIGO BARROS ROCHA

**PROPOSTA DE METODOLOGIA PARA MAPEAMENTO DE
LOCOS CONTROLADORES DE CARACTERÍSTICAS
OLIGOGÊNICAS**

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

VIÇOSA
MINAS GERAIS - BRASIL
2007

RODRIGO BARROS ROCHA

**PROPOSTA DE METODOLOGIA PARA MAPEAMENTO DE
LOCOS CONTROLADORES DE CARACTERÍSTICAS
OLIGOGÊNICAS**

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

APROVADA: 30 de março de 2007.

Prof. Cosme Damião Cruz
(Co-orientador)

Prof. Everaldo Gonçalves de Barros
(Co-orientador)

Prof. Sebastião Martins Filho

Dr. Júpiter Israel Muro Abad

Prof. Elza Fernandes de Araújo
(Orientador)

AGRADECIMENTOS

Deixo meus sinceros agradecimentos as seguintes pessoas e instituições sem as quais o presente trabalho seria impossível.

A Universidade Federal de Viçosa e ao CNPq pela oportunidade de concluir a minha formação.

A Prof. Elza Fernandes de Araújo pela dedicação e orientação constante em todas as etapas de minha formação.

Ao Prof. Cosme Damião Cruz, também orientador deste trabalho, pelo estímulo, contribuições e amizade.

Ao Prof. Everaldo Gonçalves de Barros pelas contribuições minuciosas e por ter participado da minha formação como um todo.

Ao Júpiter Israel Muro Abad e ao Prof. Sebastião Martins Filho pela contribuição na elaboração deste trabalho.

A empresa CENIBRA CELULOSE S/A pelo apoio na elaboração deste trabalho.

Aos companheiros (as) e amigos (as) do Laboratório de Genética Molecular e de Microrganismos, de Bioinformática e de curso pela convivência e por toda a contribuição recebida ao longo destes dois anos.

A todos os funcionários da Universidade Federal de Viçosa, em especial as funcionárias da secretaria de genética e melhoramento Rita e a Rose e os funcionários Evandro e Danilo pela atenção e amizade.

A minha família, aos meus pais Danilo Rocha e Maria das Graças Barros Rocha, exemplos de minha vida.

A minha esposa Roseane, pelo amor, carinho e todos os momentos de nossa vida juntos.

Aos meus filhos, Pedro e André, minha inspiração.

BIOGRAFIA

Rodrigo Barros Rocha, filho de Maria das Graças Barros Rocha e Danilo Rocha, nasceu em Viçosa - MG, em 28 de setembro de 1978.

Em 1997 formou-se no COLUNI, Viçosa - MG.

Em 1998 ingressou na Universidade Federal de Viçosa, Viçosa – MG, no curso de Ciências Biológicas, concluindo-o em Setembro de 2002.

Iniciou o curso de Mestrado em Genética e Melhoramento em Setembro de 2002, na Universidade Federal de Viçosa, defendendo tese no dia 16 de Fevereiro de 2004.

Em 2004 iniciou o curso de Doutorado em Genética e Melhoramento da UFV, defendendo sua dissertação de doutorado no dia 30 de Março de 2007.

SUMÁRIO

RESUMO	ix
ABSTRACT	x
1. INTRODUÇÃO GERAL	1
2. REVISÃO BIBLIOGRÁFICA	3
2.1. Herança oligogênica	3
2.1.1. Hipóteses genéticas e segregação ambígua	4
2.2. Marcadores moleculares e mapeamento genético	9
2.2.1. Marcadores moleculares	9
2.2.2. Mapeamento genético	10
2.2.3. Populações de mapeamento	13
2.3. Detecção de QTLs – métodos tradicionais	17
2.3.1. Método da marca simples	17
2.3.2. Método do mapeamento por intervalo simples	19
3. Pressuposições da análise de QTL e natureza das características	20
3. REFERÊNCIA BIBLIOGRÁFICAS	22
Capítulo I: Mapeamento de locos controladores de características oligogênicas de manifestação binária utilizando funções de máxima verossimilhança - Populações de retrocruzamento.....	27
1. RESUMO	28
2. ABSTRACT	29
3. INTRODUÇÃO	30

4. MATERIAL E MÉTODOS	32
4.1. Simulação de dados	32
4.1.1. Simulação do genoma	32
4.1.2. Simulação dos genitores	33
4.1.3. Simulação da característica oligogênica	34
4.2. Análises genômicas	35
4.2.1. Mapeamento genético	35
4.2.2. Análise de segregação de locos individuais – Teste de (χ^2).....	35
4.2.3. Análise de pares de marcas - estimação da porcentagem de recombinação	36
4.2.4. Análise de pares de marcas - determinação dos grupos de ligação e ordenamento das marcas	36
4.3. Comparação de genomas.....	37
4.3.1. Detecção de características oligogênicas	37
5. RESULTADOS E DISCUSSÃO	38
5.1. Simulação dos dados	38
5.2. Estimativas de probabilidade condicional	39
5.3. Funções de verossimilhança	43
5.4. Mapeamento genético	44
5.5. Detecção do OTL.....	47
5.6. Comparação entre métodos: marca simples (M.S.) e mapeamento por intervalo simples (I.S.)	52
5.6.1. Critérios para comparação entre os métodos	52
5.6.2. Marca simples e M.M.C.O.	53
5.6.3. Mapeamento por intervalo simples (I.S.) e M.M.C.O.	58

6. CONCLUSÕES	58
7. REFERÊNCIAS BIBLIOGRÁFICAS	61
Capítulo II: Mapeamento de locos controladores de características oligogênicas de manifestação binária utilizando funções de máxima verossimilhança - Populações de F2.....	
1. RESUMO	67
2. ABSTRACT	68
3. INTRODUÇÃO	69
4. MATERIAL E MÉTODOS	72
4.1. Simulação de dados	72
4.1.1. Simulação do genoma	72
4.1.2. Simulação dos genitores	72
4.1.3. Simulação da característica oligogênica	73
4.2. Análises genômicas	75
4.2.1. Mapeamento genético	75
4.2.2. Análise de segregação de locos individuais – Teste de (χ^2)....	75
4.2.3. Análise de pares de marcas - estimação da porcentagem de recombinação	76
4.2.4. Análise de Pares de Marcas - determinação dos grupos de Ligação e ordenamento das marcas	76
4.2.5. Teste de segregação ambígua	76
4.3. Comparação de genomas.....	77
4.3.1. Detecção de características oligogênicas	77
5. RESULTADOS E DISCUSSÃO	78

5.1. Simulação dos dados	78
5.2. Estimativas de probabilidade condicional	79
5.3. Funções de verossimilhança	89
5.4. Mapeamento genético	93
5.5. Detecção de OTLs	95
5.6. Segregação ambígua.....	100
5.7. Erro na declaração da hipótese genética.....	106
5.8. Estudo de herança e detecção de OTLs.....	109
6. CONCLUSÕES	110
7. REFERÊNCIAS BIBLIOGRÁFICAS	111

RESUMO

ROCHA, Rodrigo Barros, D.Sc., Universidade Federal de Viçosa, março de 2007. **Proposta de metodologia para mapeamento de locos controladores de características oligogênicas.** Orientador: Elza Fernandes de Araújo, Co-orientadores: Cosme Damião Cruz e Everaldo Gonçalves de Barros.

Características oligogênicas se caracterizam por sua herança governada por genes de maior efeito e por sua importância para as espécies vegetais cultivadas, com destaque para a resposta de resistência das plantas às doenças. A natureza discreta e a interação epistática destas características resultam em um padrão de herança que não deve ser interpretado pelos métodos tradicionais de detecção de QTL. A busca por métodos que permitam melhor interpretar a informação genética produzida durante a recombinação continua sendo um dos principais objetivos do mapeamento e identificação de QTLs. O objetivo deste trabalho é o de propor um método para mapeamento e detecção de locos controladores da expressão de características oligogênicas, OTLs ("Oligogenic Trait Loci"). Este método definido como M.M.C.O (Método de Mapeamento de Características Oligogênicas) utiliza de funções de verossimilhança para obtenção de estimativas de recombinação " r " de melhor ajuste aos dados, a partir das probabilidades condicionais de ocorrência entre o loco marcador e a herança da característica oligogênica. Os resultados mostram que o método foi adequado para detecção de OTLs em populações de retrocruzamento e em populações F2. Análises comparativas entre as metodologias mostram que as estimativas M.M.C.O. de " r " entre o loco marcador e os locos OTLs são mais precisas do que as obtidas pelos métodos de marca simples e de intervalo simples. A necessidade de conhecer previamente o padrão de herança da característica é uma das maiores limitações no estudo das características oligogênicas; e a definição não apropriada do padrão de herança resulta em uma diminuição no poder de teste da metodologia. O M.M.C.O. não necessita de informação prévia de ordenamento dos marcadores e não é influenciado pela ocorrência OTLs em um mesmo grupo de ligação. O atendimento das pressuposições para análise das características oligogênicas, a maior acurácia das estimativas, obtidas a partir da informação contida em todas as classes genotípicas, e a possibilidade de posicionar pontualmente o OTL fazem do M.M.C.O. uma estratégia de potencial para contribuir na interpretação da herança de características oligogênicas.

ABSTRACT

ROCHA, Rodrigo Barros D.Sc., Universidade Federal de Viçosa, March of 2007. **Genetic mapping of oligogenic traits using maximum likelihood functions.** Adviser: Elza Fernandes de Araújo, Co-Advisers: Cosme Damião Cruz and Everaldo Gonçalves de Barros.

Oligogenic traits are distinguished by their heritage ruled by higher effect genes and by their importance for the cultivated plants, with emphasis to the plant disease resistance heritage. The qualitative nature and epistatic interaction of these traits results in a heritage pattern that should not be interpreted through traditional methods for QTL (Quantitative Trait Loci) detection. The investigation of methods which allow a better interpretation of the genetic information produced during the recombination is still one of the major objectives of QTL detection. The objective of this work is to propose a method for detection of important loci related with the expression of oligogenic traits – OTL (Oligogenic Trait Loci). This method, defined as Oligogenic Trait Mapping Method (O.T.M.M.), uses maximum likelihood probability functions to obtain the r estimates of better adjustment to the observed data from the conditional probabilities of occurrence between the molecular markers and the OTL. The results show that the method was adequate for OTL detection in backcrossing and F₂ populations. Comparative analysis with others QTL detection methodologies shows that O.T.M.M. has the more accurate r estimates of the distance between the molecular markers and the OTL loci. The necessity of previous knowledge about the heritage pattern of the traits is one of the main limitations in the study of oligogenic traits, and the not proper definition of the correct heritage pattern results in a decrease in the accuracy of the methodology. The O.T.M.M. does not need previous information of the markers order in the linkage groups and are not influenced by the presence of two QTL in the same linkage group. The focus in the analysis presumes of the oligogenic heritage, the higher accuracy of the r estimates, obtained from the information contained in all genotypic classes, and the position of the OTL in the map, contribute to make O.T.M.M. a powerful tool for the analysis of oligogenic traits.

1. INTRODUÇÃO GERAL

O desenvolvimento do ramo das ciências biológicas relacionado com o estudo da herança e da variação dos caracteres hereditários associado às técnicas modernas de estudo da molécula do DNA tem permitido a identificação de regiões do genoma da maior importância para a expressão de quaisquer características. Associado às expectativas teóricas fundamentadas em métodos biométricos e quantitativos, o conhecimento produzido nos centros de pesquisas tem permitido não somente o entendimento dos padrões de herança como também a manipulação do processo da hereditariedade.

A experimentação nos programas de melhoramento vegetal baseia-se, principalmente, na mensuração sistemática da resposta do caráter em campo e na aplicação de conceitos da genética. A associação das técnicas modernas de análise genômica aos procedimentos clássicos de seleção é desafio contemporâneo para os pesquisadores. O desenvolvimento das técnicas de marcadores moleculares, em meados da década de 80, aumentou a utilidade prática dos mapas de ligação para a seleção de plantas e direcionamento de cruzamentos. No entanto, a incorporação destas estratégias na rotina dos programas de melhoramento tem se mostrado limitada, principalmente pelos esforços necessários para a identificação dos QTLs ("Quantitative Trait Loci"), de alta acurácia no estudo da herança de características quantitativas de distribuição contínua.

Além da relevância econômica das características quantitativas para os programas de melhoramento, as características governadas por poucos genes têm sido consideradas importantes em muitas culturas, com destaque para a resistência das plantas às doenças. A natureza discreta da resposta destas características e a interação epistática dos genes resultam em um padrão de herança que não deve ser analisado utilizando os métodos que pressupõem normalidade e herança governada por muitos genes de pequeno efeito.

Neste cenário, a busca por novos métodos e estratégias de análise é relevante, uma vez que a redução no poder de teste causada pelo não atendimento das pressuposições estatísticas compromete a interpretação dos resultados. O objetivo deste trabalho é propor uma nova abordagem para a detecção dos locos controladores da expressão de características oligogênicas

(OTLs – “Oligogenic Trait Loci”). Tal estratégia baseia-se na utilização de funções de verossimilhança visando prever novas possibilidades de segregação entre marcadores moleculares e locos controladores de características oligogênicas para a obtenção de estimativas de distância entre os marcadores e o loco OTL.

- Objetivo Geral:

- Propor e comprovar a adequação do método para mapeamento de características oligogênicas (M.M.C.O.) em populações de retrocruzamento e em populações F2.

- Objetivos Específicos:

- Apresentar a metodologia utilizada para obtenção das funções de verossimilhança;

- Comparar a metodologia apresentada com os métodos tradicionalmente utilizados para a detecção de QTLs; e

- Considerar as particularidades em relação às populações mais utilizadas no melhoramento de plantas (F2 e retrocruzamento) e em relação às interações epistáticas mais comuns.

2. REVISÃO BIBLIOGRÁFICA

2.1. Herança oligogênica

Após a redescoberta do trabalho de Mendel (MENDEL, 1866), em 1900, várias pesquisas foram realizadas visando ampliar e validar suas conclusões a respeito do mecanismo de herança e variação hereditária das características. A natureza discreta e monogênica das características analisadas por Mendel permitiu a identificação de fatores, mais tarde chamados de genes, que governam a herança das características. A identificação e a quantificação do efeito destes fatores para a expressão das características são ainda objetos de trabalhos contemporâneos que se utiliza de ferramentas de manipulação da molécula do DNA (SCHUSTER & CRUZ, 2004).

Em geral, as características de maior relevância econômica para programas de melhoramento vegetal apresentam a expressão governada por muitos genes de pequeno efeito (ZOBEL & JETT, 1995). Embora os princípios genéticos de segregação das características métricas sejam mendelianos, a identificação da segregação individual dos genes é complexa (CRUZ et al. 2004), fazendo com que a média e a variância da distribuição dos valores métricos sejam essenciais para o estudo da herança e a variação destas características (LYNCH & WALSH, 1998). A interpretação genética dos modelos biométricos pressupõe a herança governada por inúmeros genes de efeito pequeno e equivalentes para a expressão da característica (“Modelo Infinitesimal de Fisher”) (FALCONER, 1987).

Por sua vez, as características de expressão governadas por poucos genes de maior efeito têm-se mostrado importantes na condução dos programas de melhoramento de várias culturas, com destaque para a resposta de resistência de plantas a algumas doenças (ALZATE-MARIN et al. 2005; VIJAYALAKSHMI et al. 2005; ALFENAS et al. 2004; JUNGHANS et al. 2003; YU et al. 2000). As características oligogênicas possuem padrão de herança mendeliano governado por dois, três ou mais genes de interação epistática, podendo ser alocadas em classes (FALCONER, 1987).

A caracterização do padrão da herança é um dos pontos mais importantes para a manipulação dos componentes hereditários e da fração

herdável, uma vez que as estratégias de seleção para estudo genético são bastante contrastantes de acordo com a natureza da característica: se monogênica, oligogênica ou quantitativa (CRUZ et. al. 2004). Várias características apresentam padrão de herança governado por genes de maior efeito em plantas: espessura de casca (ZOBEL & JETT, 1995), acúmulo de óleos essenciais (ZOBEL & JETT, 1995) entre outras. Para o melhoramento vegetal, a resistência das plantas a patógenos é, provavelmente, a característica de natureza oligogênica mais importante (ALZATE-MARIN, et al. 2005; ALFENAS et al. 2004). A resistência ao *Ceratocystis* (ALFENAS et al. 2004) e a resistência à Ferrugem (VIJAYALAKSHMI et al. 2005; JUNGHANS et al. 2003; WILCOX et al. 1996) são exemplos de características de provável herança oligogênica.

2.1.1. Hipóteses genéticas e segregação ambígua

Um dos pontos mais importantes para o estudo das características da distribuição discreta é o teste de hipótese genética que permite, com uma margem de erro, concluir a respeito do padrão predominante da segregação: se governada por um, dois ou mais genes e o tipo de interação predominante (LIU, 1997). Neste contexto, define-se como hipótese genética o mecanismo de herança mais provável de uma característica baseado na avaliação dos dados disponíveis até o momento.

O teste de χ^2 tem se mostrado bastante prático e eficiente para o teste de hipóteses de padrões de segregação, uma vez que considera os desvios ocorridos entre valores previstos e observados e o número de observações avaliado (SCHUSTER & CRUZ, 2004). Para a execução do teste são comparados dois valores o $\chi^2_{calculado}$ e o $\chi^2_{tabelado}$, sendo o primeiro obtido a partir de dados experimentais, segundo a expressão apresentada a seguir:

$$(i) \quad \chi^2_{calculado} = \sum_i \left[\frac{(O_i - E_i)^2}{E_i} \right]$$

Em que O_i e E_i representam, respectivamente, os valores observados e

esperados para a i-ésima classe fenotípica. O χ^2_{tabelado} é obtido a partir da informação do número de graus de liberdade associado à hipótese genética testada e do nível de significância desejado. A decisão a respeito da não rejeição da hipótese genética de segregação é feita da seguinte maneira:

Se $\chi^2_{\text{calculado}} \geq \chi^2_{\text{tabelado}}$, rejeita-se H_0 .

Se $\chi^2_{\text{calculado}} < \chi^2_{\text{tabelado}}$, não se rejeita H_0 .

O teste de χ^2 é sensível ao tamanho da amostra, ou seja, quanto maior for o tamanho da amostra, menor deverá ser o desvio para a não rejeição da hipótese nula. Um número reduzido de observações pode resultar na impossibilidade da discriminação de uma ou mais hipóteses genéticas. Determinadas hipóteses genéticas podem ser de discriminação bastante dificultada, sendo necessária, em casos mais complicados, a análise de milhares de indivíduos.

LIU (1997) apresenta um estimador que permite a comparação de hipóteses genéticas em relação ao número de eventos que deveriam ser analisados para discriminá-las considerando a distribuição de χ^2 e um nível crítico associado. Como mostrado por LIU (1997), a obtenção considera a resolução do sistema de equações apresentado abaixo (ii e iii), em que x e y representam padrões de segregação diferentes, e a e b , as frequências de suas classes, como mostrado abaixo:

$$(ii) \quad \frac{(a - xb)^2}{nx}$$

$$(iii) \quad \frac{(a - yb)^2}{ny}$$

O desenvolvimento deste sistema de equações permite a obtenção do estimador do número de indivíduos necessários para discriminar duas hipóteses genéticas:

$$(iv) \quad n \geq \left[\frac{1 + \sqrt{(xy)}}{\sqrt{x} + \sqrt{y}} \right]^2 \chi^2_{\alpha,1}$$

A título de exemplificação, a determinação do número de indivíduos necessários para discriminar a hipótese de segregação da herança governada por um gene de dominância completa (3:1), da herança governada por dois genes de dominância completa (9:7), com nível de significância de 5% ($\alpha = 5\%$), está sumarizada nos dados apresentados a seguir:

$$x = \frac{3}{1}$$

$$y = \frac{9}{7}$$

$$\chi^2_{tab.5\%,1} = 3,841$$

$$\sqrt{(xy)} = 1,924$$

$$n = 94.3$$

Assim, conclui-se que 95 observações são necessárias para separar as duas hipóteses a nível de 5% de probabilidade. A impossibilidade de discriminar uma ou mais hipóteses genéticas em função do número de indivíduos avaliados é definida como segregação ambígua.

Diferentemente da abordagem métrica, que, em relação a uma série de pressuposições, suporta desvios em relação ao não atendimento, a consideração inapropriada de uma hipótese genética pode resultar em problemas de interpretação. O equívoco nas conclusões em relação à natureza da herança da característica, ao genótipo de genitores, direcionamento de cruzamentos e posicionamento dos genes nos mapas de ligação apresenta potencial para causar prejuízos na continuidade da pesquisa.

À disposição dos melhoristas existem várias configurações de cruzamentos controlados que podem ser utilizadas para não rejeitar uma hipótese genética. Os cruzamentos mais explorados nos programas de melhoramento vegetal são: cruzamentos controlados entre pais contrastantes, cruzamentos-teste, retrocruzamentos, autofecundações e também a análise de cruzamentos ao acaso. Maiores informações a respeito da importância e eficácia de cada uma destas estratégias podem ser obtidas em CRUZ et al.

(2003); CRUZ et al. (2004) e LYNCH & WALSH(1998)

CRUZ et al. (2004) apresentam uma estatística de utilidade para estimar o número de indivíduos necessários para concluir a respeito da herança de uma característica de distribuição discreta com determinado nível de confiança. Este estimador permite estimar a probabilidade de erro associada à inferência em relação à identidade dos genótipos do genitor baseando-se no genótipo de sua progênie, por meio da expressão apresentada a seguir:

$$(v) \quad \alpha = (P(A_))^n$$

em que α é o erro que se comete ao desprezar a probabilidade de um indivíduo que produza n descendentes de fenótipo dominante $P(A_)$ seja heterozigoto (Aa) e n é o número de descendentes da progênie. O complemento de α é denominado de certeza e definido como:

$$(vi) \quad c = 1 - \alpha$$

em que c é denominada de certeza, ou probabilidade de se concluir corretamente em relação à identidade do genótipo. Desta maneira, o número de indivíduos necessários para inferir a respeito da identidade de uma matriz pode ser expresso como:

$$(vii) \quad n = \frac{\text{Log}(1 - \alpha)}{\text{Log}[P(A_)]}$$

O gráfico mostrado a seguir foi construído considerando os diferentes valores de c associados aos padrões de segregação: 1:1, 3:1, 15:1, 9:7, 13:3, 63:1, 37:27. Os padrões de segregação que possuem classes mais raras apresentam menores valores de c para progênies de mesmo tamanho.

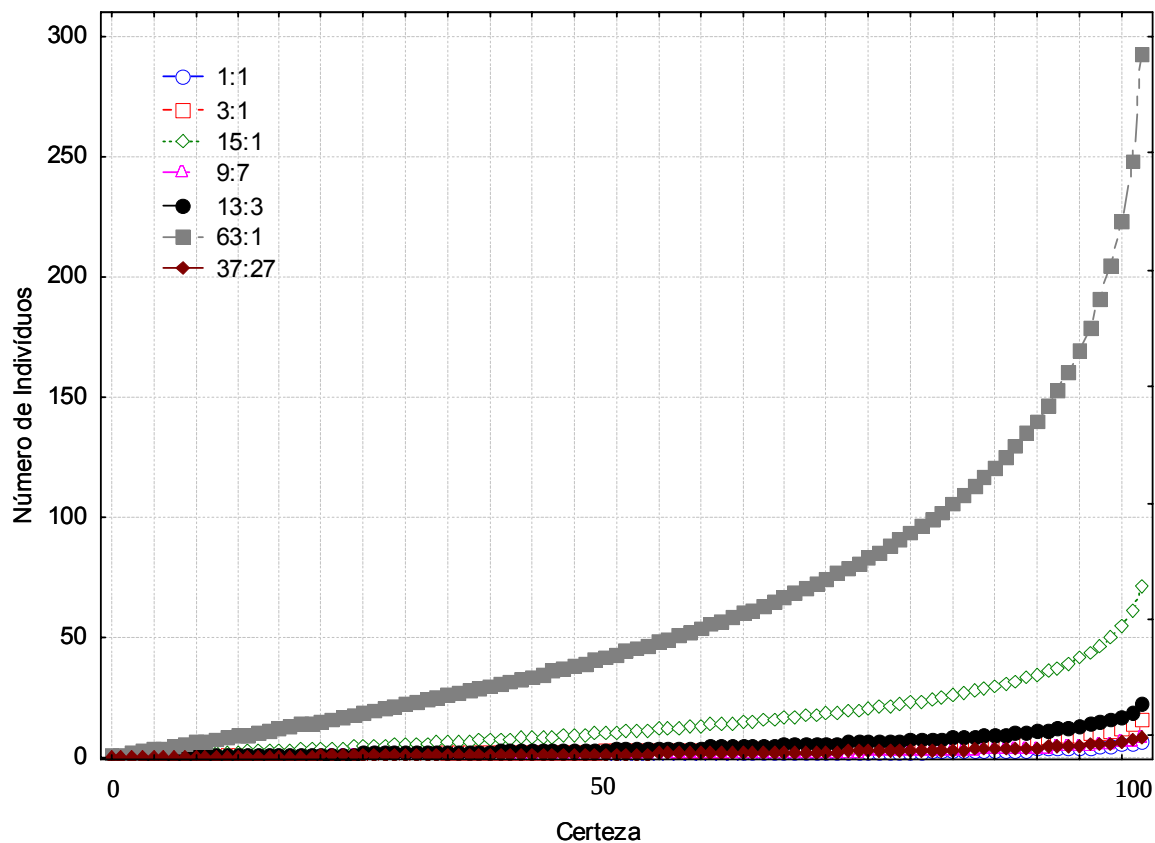


Figura 1 – Relação entre tamanho da progênie e a certeza da predição da identidade genotípica, considerando as probabilidades de ocorrência do genótipo heterozigoto para o gene marcador ($P(Aa)$), iguais, respectivamente, a 1:1, 3:4, 15:16, 9:16, 13:16, 63:64, 37:64

2.2. Marcadores moleculares e mapeamento genético

2.2.1. Marcadores moleculares

O desenvolvimento de técnicas de marcadores moleculares em meados da década de 80 disponibilizou ferramentas de grande impacto para o controle dos programas de melhoramento e seu auxílio (LIU 1997, BRONDANI et al. 1998). Os marcadores moleculares possibilitam verificar as diferenças entre dois genótipos a partir das suas seqüências de DNA, destacando-se por permitir o isolamento do efeito ambiental e a caracterização genética da população em qualquer fase do desenvolvimento (MACKAY, 2001). Em sistemas biológicos em que os indivíduos apresentam longos ciclos de vida e as características de interesse são de baixa herdabilidade, como observado em espécies florestais, o ganho genético resultante da utilização de marcadores é potencializado (YOUNG et al. 2000).

A caracterização de regiões genômicas repetitivas permitiu o desenvolvimento de marcadores com elevado conteúdo médio de informação. A repetibilidade, o multialelismo e a natureza co-dominante fazem dos marcadores microssatélites uma excelente ferramenta para estudos genéticos (BRONDANI et al. 1998). Os marcadores microssatélites são obtidos por meio da amplificação de regiões repetitivas do genoma, utilizando iniciadores específicos. Segundo CHARLESWORTH et al. (1994), uma grande parte dos genomas do DNA dos eucariotos é constituída por DNA de seqüência repetitiva. Tais regiões podem ser regiões altamente repetitivas, chamadas de DNA satélite, ou regiões moderadamente repetitivas, os microssatélites e elementos transponíveis (VARSHNEY et al. 2005). Neste contexto, as regiões microssatélites podem ser definidas como regiões repetitivas organizadas em tandem compostas por um número variável de elementos repetitivos na seqüência de um a seis pares de base, encontrado no genoma de vertebrados, insetos e plantas (SCHLOTTERER et al. 2000). Os microssatélites são considerados regiões instáveis do genoma comparadas às regiões de DNA não repetitivo, pois estão mais sujeitas à ação de alterações do tipo adição ou deleção de repetição (MORGANTE et al. 2002; SCHLOTTERER et al. 2000). A taxa de alterações nestas regiões pode ser tão alta quanto uma alteração a cada 1000 pares de base. Esta instabilidade inerente às regiões repetitivas

resulta em grande número de locos polimórficos intraespecíficos, que tornam estas regiões bastante adequadas para estudos genéticos entre indivíduos de uma mesma espécie (VARSHNEY et al. 2005; ROCHA et al. 2002; BRONDANI et al. 1998).

A variabilidade entre estas seqüências pode ser avaliada por PCR utilizando iniciadores que flanqueiam estas regiões. A necessidade de desenvolvimento de oligonucleotídeos específicos representa a principal limitação desta metodologia, tendo em vista a considerável quantidade de recursos necessários para o desenvolvimento destes oligonucleotídeos de aplicação limitada à espécie ou ao gênero do organismo para o qual foram desenvolvidos (MORGANTE et al. 2002).

Marcadores moleculares podem ser utilizados com diferentes finalidades em estudos de genética: proteção de cultivares (BRONDANI et al. 1998), confirmação de cruzamentos controlados e identidade clonal (ROCHA et al. 2002), seleção de genitores (DIAS et al. 2001), estudo de diversidade genética (MURO ABAD et al. 2005a; MURO ABAD et al. 2005b; DIAS et al. 2005), mapeamento genético (GRATTAPAGLIA & SEDEROFF, 1994) e detecção de QTLs (VERHAEGEN et al. 1997).

2.2.2 Mapeamento genético

A utilização de técnicas de marcadores moleculares no estudo de características quantitativas permite identificar maiores frações da variância genética total relacionadas com regiões específicas do genoma (SCHUSTER & CRUZ, 2004; LIU, 1997). O termo QTL (“Quantitative trait loci”) foi apresentado pela primeira vez por GELDERMANN (1975) e se baseia na pressuposição de existência de locos de maior importância para a expressão de determinada característica métrica. Além da determinação da posição do QTL ao longo do grupo de ligação, estudos de mapeamento permitem a obtenção de estimativas de ação gênica: efeitos aditivos e de interação entre genes e alelos (SCHUSTER & CRUZ, 2004; LIU, 1997).

Historicamente, mapas de ligação foram pouco utilizados no auxílio à seleção em programas de melhoramento (BOREVITZ & CHORY, 2006). A

utilização de marcadores fenotípicos e citológicos e a falta de mapas integrados que relacionem vários tipos de marcadores são as principais dificuldades para a utilização de mapas genéticos em programas de melhoramento (DEKKERS & HOSPITAL, 2001). O número reduzido de marcadores citológicos e fenotípicos faz com que a probabilidade de encontrar um marcador relacionado com características de interesse seja reduzida, dificultando o mapeamento genético (WEIR, 1996).

O objetivo da utilização de marcadores moleculares é incrementar o ganho genético por unidade de tempo (BADU et al.2004; MURO ABAD et al. 2000). Neste sentido, a construção de mapas de ligação utilizando marcadores moleculares é a estratégia que vem sendo utilizada para a detecção de locos relacionados com a expressão de características quantitativas. A disponibilidade de marcadores genéticos neutros cuja herança pode ser acompanhada sem influência do ambiente permitiu a construção de mapas de ligação para várias espécies vegetais (MACKAY, 2001).

Em geral, o primeiro passo na construção de um mapa de ligação refere-se à escolha dos cruzamentos de forma a maximizar o polimorfismo genético (GRATTAPAGLIA & SEDEROFF, 1994; VERHAEGEN et al. 1997). Uma vez selecionados os genitores, a etapa seguinte é o desenvolvimento de uma progênie segregante composta de centenas de indivíduos. Tradicionalmente a população segregante é derivada do retrocruzamento ou da autofecundação da geração F1 proveniente do cruzamento de dois genitores selecionados (SCHUSTER & CRUZ, 2004; ROCHA et al. 2003). A partir da análise da segregação, as estimativas de recombinação são obtidas para posicionamento e ordenamento dos marcadores nos grupos de ligação (CRUZ & SILVA, 2006).

O método da máxima verossimilhança é utilizado por vários pesquisadores para obtenção das estimativas de recombinação (SCHUSTER & CRUZ, 2004; LIU, 1997; LYNCH & WALSH; 1998, WEIR, 1996). A máxima verossimilhança permite a obtenção de estimativas de variância mínima que são estimadas a partir da informação contida em todas as classes genotípicas. Esta propriedade é de grande relevância, principalmente se comparada com métodos mais simples de estimação das distâncias de recombinação tais como a de simples contagem da proporção dos indivíduos recombinantes em relação

ao total (SCHUSTER & CRUZ, 2004). Para a utilização da função de verossimilhança é necessário o conhecimento da função de densidade de probabilidade da distribuição de uma variável aleatória x , sendo que a probabilidade de ocorrência das classes genotípicas nas populações segregantes avaliadas neste estudo assumem distribuição multinomial apresentada a seguir, segundo notação de SCHUSTER & CRUZ, (2004):

$$(viii) \quad L(p_i / n_i) = \lambda p_1^{n_1} p_2^{n_2} p_3^{n_3} \dots p_n^{n_n}$$

Em que :

$$(ix) \quad \lambda = \frac{N!}{n_1! n_2! \dots n_n!}$$

$$(x) \quad N = \sum_{i=1}^n n_i, \text{ e}$$

N é o tamanho da população

n_i é o número de ocorrências do evento x_i com probabilidade p_i .

Considerando que os máximos da função de verossimilhança não se alteram com a aplicação do logaritmo natural, em geral, ele é aplicado nos termos da função anterior visando a uma incorporação de uma facilidade algébrica no processo de derivação e obtenção do melhor estimador,

$$(xi) \quad \ell(p_i / n_i) = \ln(\lambda p_1^{n_1} p_2^{n_2} p_3^{n_3} \dots p_n^{n_n}),$$

que também pode ser escrito como:

$$(xii) \quad \ell(p_i / n_i) = \ln \lambda + \ln p_1^{n_1} + \ln p_2^{n_2} + \ln p_3^{n_3} \dots + \ln p_n^{n_n}$$

A primeira derivada da função suporte em relação ao parâmetro que se deseja estimar permite a obtenção de uma segunda função, denominada função score (f'). A função score é utilizada para obter o estimador de máxima verossimilhança relacionado com o ponto de máximo da distribuição:

$$(xiii) \quad f'[\ell(p_i / n_i)] = \frac{\partial \ell(p / n)}{\partial p}$$

Considerando que, neste caso, o único parâmetro a ser estimado é a frequência de recombinação (r) entre dois marcadores adjacentes, a frequência de recombinação pode ser obtida por meio da igualdade da função a 0, conforme mostrado a seguir.

$$(xiv) \quad f'[\ell(p_i / n_i)] = \frac{\partial \ell(p / n)}{\partial p} = 0$$

2.2.3. Populações de mapeamento

A diversidade do sistema reprodutivo das espécies vegetais permite a utilização de diferentes populações experimentais e delineamentos genéticos para a construção de mapas de ligação. Algumas culturas, tais como o feijão ou a soja, por exemplo, possibilitam a obtenção de vários tipos de população para mapeamento: populações F2 (BUNYAMIN et al. 2002), populações de retrocruzamento (MATTHEWS et al. 2001) e populações de “RIL” (“Recombinant Inbreed Lines”) (BOUCHEZ et al. 2002). Cada população ou progênie possui propriedades genéticas distintas que devem ser consideradas pelo pesquisador durante o planejamento da pesquisa (ROCHA et al. 2004).

Além das questões práticas, a natureza e a acurácia da informação genética desejadas devem ser consideradas pelo pesquisador na escolha da população segregante para o mapeamento genético. A variância e o índice de informatividade das estimativas de recombinação de máxima verossimilhança permitem avaliar a acurácia das estimativas de recombinação (WEIR 1996, LIU 1997). Para um número elevado de observações, a variância de um estimador de máxima verossimilhança pode ser definida como (WEIR 1996) :

$$(xvi) \quad V(\theta) \cong \frac{1}{E[I(\theta)]}$$

O índice de informatividade, definido com o negativo da segunda derivada da função de verossimilhança, varia conforme o tipo de população e a

saturação média do mapa (SCHUSTER & CRUZ, 2004). A Figura 2 mostra os valores da variância e do índice de informatividade para diferentes tipos de população:

$$(xvi) \quad I(\theta) = -\left(\frac{\partial^2 L(\theta; x)}{\partial \theta} \right)$$

Entre as populações tradicionalmente avaliadas no melhoramento de plantas, as populações F2 derivadas do F1 (HARUSHIMA et al. 1998) e as populações de retrocruzamento (DARVASI & SOLLER 1995; WEIGUO et al. 2002; HE et al. 2001) são intensivamente utilizadas para o mapeamento por sua praticidade e facilidade de obtenção (SCHUSTER & CRUZ, 2004).

Populações F2 produzidas pela autofecundação da progênie F1, derivada de cruzamento entre genitores divergentes, são uma das preferidas por muitos pesquisadores para o desenvolvimento de mapas genéticos (HARUSHIMA et al. 1998) A sua facilidade de obtenção, a alta informatividade (produz duas vezes mais marcadores informativos do que uma população de retrocruzamento) e a possibilidade de avaliação dos efeitos de natureza aditiva e dominante em separado fazem da população F2 mapeada com marcadores codominantes uma das configurações genéticas de maior precisão das estimativas de recombinação. A maior dificuldade da multiplicação clonal de todos os indivíduos de população F2 é provavelmente uma das principais limitações desta população (LIU, 1997).

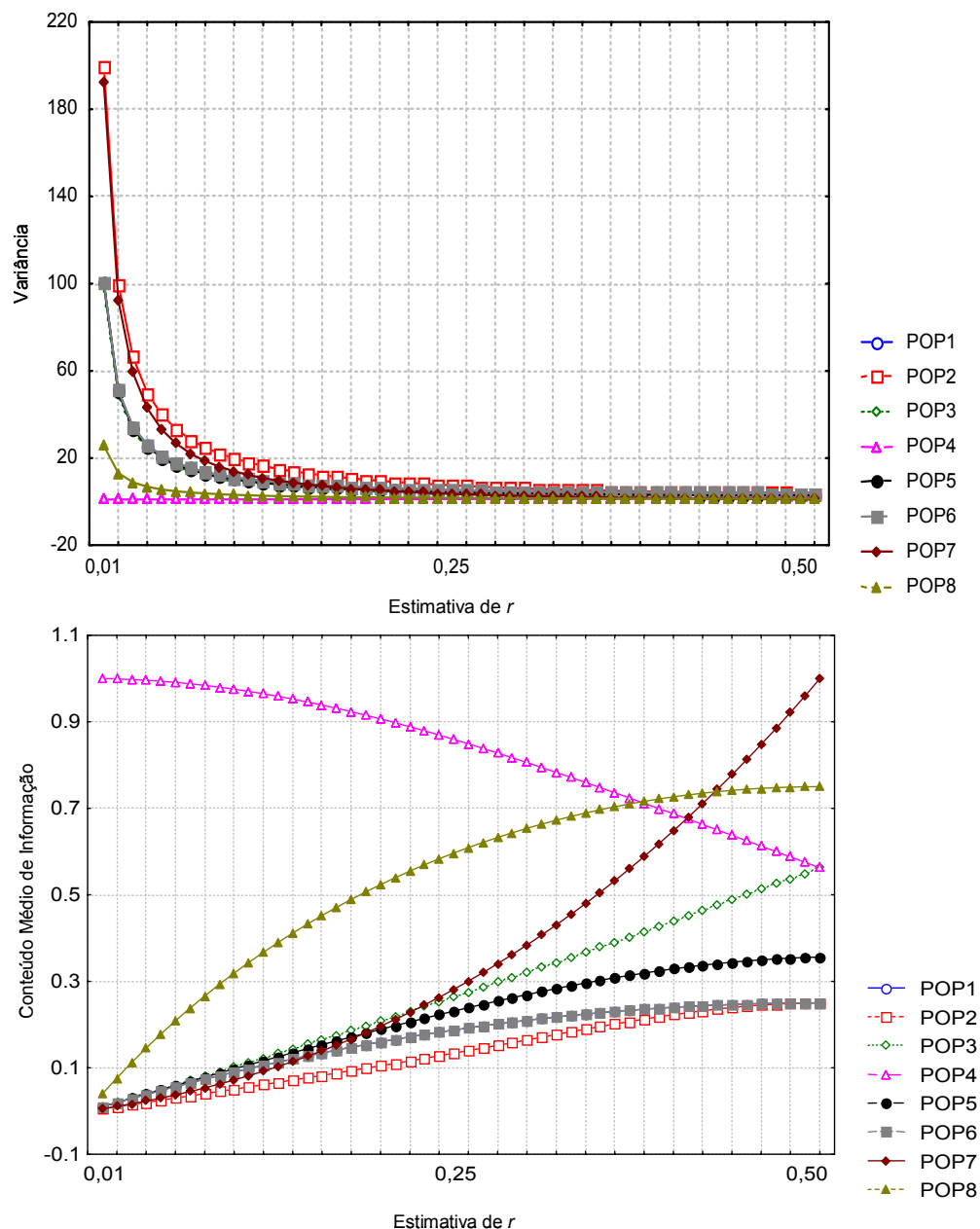


Figura 2- Distribuição dos valores de variância e do conteúdo médio de informação das estimativas de “r” para 8 tipos de populações: POP1- Retrocruzamento mapeado com marcador dominante, POP2- F2 mapeada com marcador codominante, POP3- F2 mapeada com marcador dominante (aproximação), POP4- F2 mapeada com marcador dominante (repulsão), POP5- F2 mapeada com marcadores dominante e co-dominantes, POP6- Duplo haplóide, POP7- RIL, POP8- F1 mapeada com marcador dominante (segregação misturada 1:1, 3:1). (Figura retirada de Rocha et al. 2004)

As populações de retrocruzamento apresentam apenas duas classes genotípicas segregantes. A rapidez na obtenção da população e o fato de que a segregação obtida representa a composição dos grãos de pólen do progenitor doador são as principais particularidades desta população. O menor conteúdo de informação, a impossibilidade de discriminar o componente genético de dominância do componente aditivo e a dificuldade de obtenção de repetições dos indivíduos são suas principais limitações (SCHUSTER & CRUZ 2004).

O principal benefício da distribuição dos valores do conteúdo médio de informação e da variância associada às frequências de recombinação é auxiliar na escolha da população de mapeamento mais apropriada, considerando também outras limitações experimentais. A Figura 2 mostra a existência de considerável diferença nos valores destas estimativas que influenciam na qualidade dos mapas de ligação obtidos a partir de um número limitado de indivíduos.

Para ilustrar a variabilidade inerente aos trabalhos de mapeamento, a Tabela 1 mostra uma síntese de alguns trabalhos de referência de mapeamento e detecção de QTL em espécies vegetais. Em geral, consideram-se, com alguma flexibilidade, 200 indivíduos e saturação média de 5 cM como medidas de referência para o desenvolvimento de mapas de qualidade (SCHUSTER & CRUZ, 2004).

Tabela 1- Sumário de alguns trabalhos de mapeamento e detecção de QTLs em plantas, destacando-se o tamanho da população experimental, a técnica de marcadores moleculares utilizada e a saturação média do mapa.

Referência	Espécie	Tipo de população	Nº Ind.	Marcadores	Saturação
Missiagia, 2005	Eucalipto	F1 (pseudotestcross)	270	SSR (155)	6.6 cM
Novaes, 2005	Eucalipto	F1 (pseudotestcross)	188	SSR (235 locos)	4.2 cM
Mccouch, et al. 2002	Arroz	F2 e retrocruzamento	295	SSR (2240 locos)	menor que 1 cM
Meksem et al. 2001	Soja	RIL	100	SSR (133)	20 cM
Cardinal et al. 2001	Milho	RIL	183	120 RFLP e 65 SSR	9 cM
Nodari et al. 1993	Feijão	F _{2:3}	70	SSR (152 locos)	9 cM
Yin, 2001	<i>Populus</i> spp.	Retrocruzamento	180	SSR (544)	menor que 1 cM
Cervera et al. 2001	<i>Populus deltoides</i>	F1 (pseudotestcross)	232	390 AFLP e 63 SSR	10 cM
Cervera et al. 2001	<i>Populus nigra</i>	F1 (pseudotestcross)	105	355 AFLP e 49 SSR	12.5 cM
Cervera et al. 2001	<i>Populus trigocarpa</i>	F1 (pseudotestcross)	127	287 AFLP e 76 SSR	11.2 cM

2.3. Detecção de QTLs – métodos tradicionais

O desenvolvimento da teoria de identificação de QTLs (“Quantitative trait loci”), por associação a marcadores moleculares, tem permitido a caracterização de marcadores neutros de herança mendeliana, diretamente correlacionados com a variação genética de características quantitativas. O monitoramento da herança complexa destas características por meio de marcas de herança simples é uma das aplicações mais imediatas do desenvolvimento desta tecnologia (DEKKERS & HOSPITAL, 2001; MOREAU et al. 2000; LANDE & THOMPSON, 1990). Esta abordagem baseia-se na pressuposição da existência de locos de maior efeito na expressão das características métricas GELDERMANN, (1975).

Estratégias utilizadas para a detecção e caracterização dos QTLs dependem principalmente do mecanismo de herança da característica, da população avaliada, do número de indivíduos genotipados e do tipo de marcador molecular (BOREVITZ & CHORY, 2006; DEKKERS & HOSPITAL, 2001). Além destas estratégias, destacam-se também o efeito do QTL na expressão da característica, a precisão experimental, a herdabilidade da característica avaliada e os efeitos da dominância (LIU, 1997).

Do método mais simples para o mais complexo, os métodos mais utilizados para detecção de QTLs em populações provenientes de cruzamento controlado são: mapeamento por marca simples, mapeamento por intervalo simples e mapeamento por intervalo múltiplo (SCHUSTER & CRUZ, 2004). Os métodos divididos em relação aos que consideram a avaliação de apenas uma única marca por vez são: método da marca simples ou na avaliação de duas marcas simultaneamente e métodos de mapeamento por intervalo. Até o momento, a análise baseada em uma única marca é considerada exploratória em comparação com a precisão das estimativas obtidas com métodos de intervalo.

2.3.1- Método da marca simples

O teste da marca simples testa a associação entre genótipos que possuem uma forma alélica específica utilizando o teste t e a análise de

variância ou regressão, visando testar a significância da diferença entre grupos que apresentam ou não determinada marca molecular. Considerando duas formas alélicas A_1 e A_2 e valores de médias respectivamente μ_1 e μ_2 , a hipótese nula testada pelo teste t é definida como $H_0 : \mu_1 = \mu_2$ contra a hipótese alternativa $H_a : \mu_1 \neq \mu_2$. A estatística do teste é mostrada por SCHUSTER & CRUZ(2004):

$$(xvi) \quad t = \frac{\hat{\mu}_1 - \hat{\mu}_2}{\sqrt{\hat{V}(\hat{\mu}_1 - \hat{\mu}_2)}} , \text{ associado a } n_1 + n_2 - 2 \text{ graus de}$$

liberdade.

$$(xvii) \quad \hat{V}(\hat{\mu}_1 - \hat{\mu}_2) = \frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}$$

S_1^2 e S_2^2 são as variâncias obtidas para a característica quantitativa entre os indivíduos que possuem o marcador molecular e os que não o possuem.

No caso da regressão, a hipótese nula testada é $H_0 : \beta_1 = 0$ contra a hipótese alternativa $H_a : \beta_1 \neq 0$, utilizando o modelo:

$$(xviii) \quad Y_j = \beta_0 + \beta_1 X_j + \varepsilon_j$$

em que:

Y_j : valor da característica quantitativa no i-ésimo indivíduo;

X_j : código do marcador (1 para presença e 0 para ausência);

β_0 : intercepto da regressão (média da característica);

β_1 : inclinação da reta; e

ε_j : erro aleatório na j-ésima observação em que $\varepsilon_j \sim N(0, \sigma^2)$.

No caso específico da população de retrocruzamento, existe uma equivalência no uso do teste t e da regressão, uma vez que, geneticamente, os dois testes equivalem a testar $(1-2r)(a-d)$ em que r é a distância em unidades de recombinação entre o marcador testado e o QTL, e a e d são, respectivamente, os efeitos aditivos e de dominância do QTL. Neste caso, a significância do teste da marca simples é resultado da ação confundida da distância entre a marca e o QTL, do efeito aditivo e do efeito de dominância do QTL, não sendo possível diferenciar se o efeito significativo pelo teste se deve

à presença de um QTL de maior efeito ou se é devido à proximidade do marcador a um QTL de menor efeito. Em linhas gerais, além do confundimento dos efeitos, considera-se que o teste da marca simples possui pequeno poder de detecção de QTL, sendo utilizado, na maioria dos casos, como uma análise exploratória dos dados. No entanto, por não necessitar da informação prévia de ordenamento das marcas ao longo dos grupos de ligação, pode ser uma análise exploratória útil quando esta informação não estiver disponível (LANDER & BOTSTEIN, 1989).

2.3.2- Método do mapeamento por intervalo simples

O método de mapeamento por intervalo apresentado por LANDER & BOTSTEIN (1989) é considerado mais preciso que o método da marca simples, pois se baseia na identificação de QTLs por meio da análise dos intervalos entre marcadores adjacentes ao longo de todos os grupos de ligação. Os testes do mapeamento por intervalo são feitos pela incorporação da informação dos marcadores que flanqueiam determinado intervalo no modelo de regressão. A título de exemplificação, a seguir é apresentado o modelo de mapeamento por intervalo da população de retrocruzamento, segundo SCHUSTER & CRUZ(2004):

$$Y_j = \gamma + gx_j^* + \varepsilon_j$$

Em que:

Y_j é o valor mensurado para a característica no i-ésimo indivíduo;

$\gamma = u + d_j$: Valor genético do efeito $u+d$ da característica na população;

$g = a + d$: Efeito confundido aditivo e devido à dominância do loco em estudo;

x_j^* Variável condicionadora cujos valores são dependentes dos genótipos dos marcadores que flanqueiam o QTL do indivíduo j ; e

ε_j erro aleatório na j -ésima observação em que $\varepsilon_j \sim N(0, \sigma^2)$.

De posse das probabilidades condicionais de ocorrência do genótipo

marcador e do genótipo QTL, as mesmas estimativas também podem ser obtidas pelo método da máxima verossimilhança. Neste caso específico, a consistência entre as estimativas faz destas duas, estratégias equivalentes para a obtenção das estimativas, sendo que a facilidade aritmética de obtenção das estimativas, utilizando o modelo de regressão, é um atrativo deste método (SCHUSTER & CRUZ, 2004). As estimativas de efeito e a posição dos QTLs obtidas pelo mapeamento por intervalo simples são mais precisas e acuradas do que aquelas obtidas pelo método de marcas simples. No entanto, o mapeamento por intervalo simples também possui limitações. A obtenção de estimativas viesadas de r causadas pela ocorrência de dois QTLs em um mesmo grupo de ligação é uma limitação já bem caracterizada desta metodologia. O QTL de efeito e posicionamento viesado, denominado QTL fantasma, é um artefato da técnica geralmente encontrado entre dois QTLs adjacentes em um mesmo grupo de ligação.

3- Pressuposições da análise de QTL e natureza das características

A importância das características quantitativas associada ao anseio de monitorar os fatores que governam a herança complexa motiva o desenvolvimento de novos métodos e estratégias de análise. A significativa evolução dos métodos de mapeamento tem contribuído para o estudo da genética das características métricas.

As metodologias apresentadas anteriormente apresentam pressuposições em comum, condições estabelecidas a priori que devem ser atendidas para garantir a confiabilidade do método e a precisão das estimativas. Condições estas compatíveis com a natureza de características métricas tais como: 1- aditividade dos efeitos, 2 - homogeneidade do resíduo, 3-normalidade da distribuição dos caracteres. Alguns trabalhos relatam que o atendimento parcial destas características pode não resultar em significativas reduções no poder de teste dos métodos. No entanto, espera-se que a acentuada violação das pressuposições da análise aumente a ocorrência de erros do tipo I e tipo II (SILVA, 2005).

A normalidade das características é uma das pressuposições mais comuns de não serem atendidas ou mesmo de serem atendidas apenas

parcialmente, e que não apresentam solução trivial (LIU, 1997). Alternativas como a transformação de dados ou a troca de uma distribuição por outra de melhor ajuste juntamente com derivações apropriadas das metodologias são questionáveis, caso da transformação de dados, ou de difícil implementação. Em geral, o estudo da normalidade das características quantitativas é feito objetivando inferir a precisão das metodologias de mapeamento de QTLs.

Os testes mais utilizados para verificar a normalidade de uma amostra de dados são os testes χ^2 , de Kolmogorov-Smirnov e de Lilliefors. O teste de Lilliefors apresenta vantagens por não necessitar do conhecimento prévio dos valores de média e variância que são estimados da própria amostra e pode ser aplicado sem restrição para pequenas amostras. Por considerar os dados individualmente, não perdendo informação devido a agrupamentos, o teste de Lilliefors é amplamente utilizado para avaliar sua normalidade.

A procura por novas estratégias para análise justifica-se quando o não atendimento das pressuposições básicas resulta em uma redução no poder do teste e, portanto, em estimativas de menor confiabilidade. A distribuição discreta, o variável número de genes e as interações epistáticas das características oligogênicas motivam a utilização de métodos apropriados para a natureza destas características.

3. REFERÊNCIAS BIBLIOGRÁFICAS

- ALFENAS A.C., ZAUZA, E.A.V., MAFIA R.G., ASSIS T.F. **Clonagem e doenças do Eucalipto**. Viçosa, MG. Editora UFV, 2004. 442p.
- ALZATE-MARIN A.L., CERVIGNI G.D.L., MOREIRA M.A., BARROS E.G. Seleção assistida por marcadores moleculares visando ao desenvolvimento de plantas resistentes a doenças, com ênfase em feijoeiro e soja. **Fitopatologia Brasileira**, v.30, n.4, p.333-342, 2005.
- BADU R., SUDHA K.N., PRASANNA B.M., GUPTA H.S. Integrating marker assisted selection in crop breeding – Prospects and challenges. **Current Science**, v.87, n.5, p.607-619, 2004.
- BOREVITZ, J.O., CHORY, J. Genomics tools for QTL analysis and gene discovery. **Current opinion Plant Biology**, v.7, p.132-136, 2006.
- BOUCHEZ A., HOSPITAL F., CAUSSE M., GALLAIS A., CHARCOSSET A. Marker-assisted introgression of favorable alleles at quantitative trait loci between maize elite lines. **Genetics**, v.162, p.1945-1959, 2002.
- BRONDANI, R.P.V., BRONDANI C., TARCHINI R., GRATTAPAGLIA D. Development, characterization and mapping of microsatellite markers in *Eucalyptus grandis* and *E. urophylla* . **Theoretical and Applied Genetics**, v.97, p.816-827, 1998.
- BUNYAMIN T., THOMAS E.M., PAULS K.P. Genetic mapping of agronomic traits in common bean. **Crop Science**, v.42, p.544-556, 2002.
- 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.
- CERVERA M.T., STORME V., IVENS B., GUSMÃO J., LIU B.H., HOSTYN V., SLYCKEN J.V., MONTAGU M.V., BOERJAN W. Dense genetic linkage maps of three populus species (*Populus deltoides*, *P. nigra* and *P. trichocarpa*) based on AFLP and microsatellite markers. **Genetics**, v.158 p.787-809, 2001.
- CHARLESWORTH, B., SNIEGOWSKI, P., STEPHAN, W. The evolutionary dynamics of repetitive DNA in eukaryotes. **Nature**, v.371, p.215 – 220, 1994.
- CRUZ C.D., REGAZZI A.J., CARNEIRO P.C.S. **Modelos Biométricos Aplicados ao Melhoramento Genético - Volume 1**, Viçosa , MG. Editora UFV, 2004. 480p.
- CRUZ C.D., REGAZZI A.J., CARNEIRO P.C.S., **Modelos Biométricos Aplicados ao Melhoramento Genético - Volume 2**, Viçosa , MG. Editora UFV, 2003. 585p.

- CRUZ C.D., SILVA L. C. **Análise de marcadores moleculares**. In: BORÉM A., Marcadores moleculares, Viçosa , MG. Editora UFV, 2006. 374p
- DARVASI A, SOLLER M. Advanced intercross lines, an experimental population for fine genetic mapping. **Genetics**, v.141, p.1199-1207, 1995.
- DEKKERS J.C.M., HOSPITAL F. The use of molecular genetics in the improvement of agricultural populations. **Nature reviews**, v.3, p.22-32, 2001.
- DIAS L.A.S., ROCHA, R.B., MURO ABAD J.I., SALOMÃO T.F., ALFENAS A.C. Cocoa cultivar distinctiveness and hybrid prediction using RAPD markers. **Ingenic Newsletter**, v.6, p.19–22, 2001.
- DIAS, L.A.S., ROCHA, R. B., PICOLI, E.A.T Distinctness of cacao cultivars using yield component data and RAPD markers. **Crop breeding and Applied Biotechnology**, v.5, p.47 - 54, 2005.
- FALCONER, D.S. **Introdução à genética quantitativa**. Trad. De Martinho de Almeida e Silva e José Carlos Silva. Viçosa, MG: UFV, 1987. 279p.
- GELDERMANN, H. Investigations on inheritance of quantitative characters in animals by genes markers. **Theoretical and Applied Genetics**, v.46, p.310-330, 1975.
- GRATTAPAGLIA, D., SEDEROFF, R. Genetic linkage maps of Eucalyptus grandis and E. urophylla using a pseudo-testcross mapping strategy and RAPD markers. **Genetics**, v.137, p.1121-1137, 1994.
- HARUSHIMA Y., YANO M., SHOMURA A., SATO M., SHIMANO T., KUBOKI, Y., YAMAMOTO T., LIN S.Y., BALTAZAR A.A., PARCO A., KAJIYA H., HUANG N., YAMAMOTO K., NAGAMURA Y., KURATA N., KHUSH G.S., SASAKI T.A. high-density rice genetic linkage map with 2275 markers using a single F2 population. **Genetics**, v.148, p.479-494, 1998.
- HE P., LI J.Z., ZHENG X.W., SHEN L.S., LU C.F., CHEN Y., ZHU L.H. Comparison of molecular linkage maps and agronomic trait loci between DH and RIL populations derived from the same rice cross. **Crop Science**, v.41, p.1240-1246, 2001.
- JUNGHANS D. T., ALFENAS A. C., BROMMONSHENKEL S. H., ODA S., MELLO E. J., GRATTAPAGLIA D. Resistance to rust (Puccinia psidii Winter) in Eucalyptus mode of inheritance and mapping of a major gene with RAPD markers. **Theoretical and Applied Genetics**. v.108, p.175-180, 2003.
- LANDE R., THOMPSON R. Efficiency of Marker-Assisted Selection in the Improvement of Quantitative Traits. **Genetics**, v.124, p.743-756, 1990.

- LANDER E. S., BOTSTEIN D., Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. **Genetics**, v.121, n.1, p.185-199, 1989
- LIU B.H., **Statistical Genomics: linkage mapping and QTL analysis**. Boca raton, Florida, USA: CRC Press, 1997, 610p.
- LYNCH, M.; WALSH, B. **Genetics and analysis of quantitative traits**. Sunderland, USA: Sinauer Associates, Inc.Ed., 1998, 978p.
- MACKAY T.F.C., Genetic architecture of quantitative traits. **Annual Review of Genetics**, v.35, p.303-339, 2001.
- MATTHEWS F.B., DEVINE T.E., WEISEMANN J.M., BEARD H.S., LEWERS K.S., MACDONALD M.H., PARK Y., MAITI R., LIN J., KUO J., PEDRONI M.J., CREGAN P.B. , SAUNDERS J.A. Incorporation of sequenced cDNA and genomic markers into the soybean genetic map. **Crop Science**, v.41, p. 516-521, 2001.
- MCCOUCH, S. R. TEYTELMAN L. XU Y., LOBOS K.B., CLARE K., WALTON M., FU B., MAGHIRANG R., LI Z., XING Y, ZHANG Q., KONO I., YANO M., FJELLSTROM R., DECLERCK G., SCHNEIDER D., CARTINHO S., WARE D., STEIN L. Development and Mapping of 2240 New SSR Markers for Rice (*Oryza sativa* L.). **DNA Research**, v.9, p.199-207, 2002.
- MEKSEM, K., NJITI, V.N., BANZ, W.J., IQBAL, M.J., KASSEM, M.M., HYTEN, D.L.,YUANG, J., WINTERS, T.A., LIGHTFOOT, D.A. Genomic regions that underlie soybean seed isoflavone content. **Journal of Biomedicine and Biotechnology**, v.1, n.1, p.38-44, 2001.
- MENDEL G., Experiments in plant hibridization, 1866 (Traduzido para o inglês por William Bateson, 1901). **Disponível em** <http://www.esp.org/foundations/genetics/classical/gm-65.pdf>. Acessado 27/06/06.
- MISSIAGIA A. A., **Mapeamento Genético de QTL para qualidade da madeira e florescimento precoce e estudos de expressão gênica alelo específica em *Eucalyptus* spp.** Piracicaba, 2005. 236p. Dissertação (Doutorado), Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo.
- MOREAU, L., LEMARIE, S., CHARCOSSET, A., GALLAIS, A. Economic efficiency on one cycle of marker-assisted selection. **Crop Science**, v.40, p.329-337, 2000.
- MORGANTE M., HANAFEY M., POWELL W. Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. **Nature Genetics**, v.30, p.194-200, 2002.

- MURO ABAD J.I., GOMES E.A., CÂNCIO O.N., ARAÚJO E. F. Genetic analysis of *Eucalyptus urophylla* and *E. grandis* clones selected in commercial crops from the Brazilian amazon by RAPD markers. **Silvae Genetica**, v.50, p.177-181, 2000.
- MURO ABAD, J.I., ROCHA, R.B., CRUZ, C.D., ARAÚJO, E.F. Obtainment of *Eucalyptus* spp. hybrids aided by molecular markers - SSR analysis. **Scientia Forestalis**, v.67, p.53 - 63, 2005a.
- MURO ABAD, J.I., ROCHA, R.B., CRUZ, C.D., ARAÚJO, E. F. Crosses recommendation method for obtaining *Eucalyptus* spp. hybrids assisted by molecular markers. **Crop Breeding and Applied Biotechnology**, v.5, p.71 - 73, 2005b.
- NODARI R.O., TSAI S. M, GUZMAN P., GILBERTSONT R. L., GEPTS P. Toward an Integrated Linkage Map of Common Bean. 111. Mapping Genetic Factors Controlling Host-Bacteria Interactions. **Genetics**, v.134. p.341-350, 1993.
- NOVAES E. **Mapeamento de QTLs para qualidade da madeira em *Eucalyptus grandis* x *E. urophylla* e ancoragem de clones BAC no mapa genético.** Viçosa, 2006. 171p. Dissertação (Mestrado), Universidade Federal de Viçosa.
- ROCHA R.B., MURO ABAD, J.I., ELEOTÉRIO I.P., ARAÚJO E.F. Fingerprint and genetic diversity analysis of *Eucalyptus* spp. genotypes using RAPD and SSR markers, **Scientia Forestalis** v.62, p.24-31, 2002.
- ROCHA R.B., PEREIRA J.F., CRUZ C.D., QUEIROZ M.V., ARAÚJO E.F. O mapeamento genético no melhoramento de plantas. **Biotechnologia Ciência e Desenvolvimento**. v. 30, p.27-32, 2003.
- ROCHA, R.B., CRUZ, C.D., BARROS, W. S., FERREIRA, F.M., ARAÚJO, E.F. Comparisons of segregating populations for genetic mapping. **Crop Breeding and Applied Biotechnology**., v.4, p.408 - 415, 2004.
- SILVA, L. C. **Simulação do tamanho da população e da saturação do genoma para mapeamento genético de RILs.** Viçosa, 2005. 120p. Dissertação (Mestrado), Universidade Federal de Viçosa.
- SCHLOTTERER, C. Evolutionary dynamics of microsatellite DNA. **Chromosoma**, v.109, p.365-371, 2000.
- SCHUSTER I., CRUZ C.D. **Estatística genômica aplicada a populações derivadas de cruzamentos.** Viçosa , MG. Editora UFV, 2004. 585p.
- VARSHNEY R.K., GRANER A., SORRELLS M.E., Genic microsatellite markers in plants: features and applications. **Trends in Biotechnology**, v.23, n.1, p.48-55, 2005.

- VERHAEGEN, D., PLOMION, C., GION, J.-M.; POITEL, M., COSTA, P., KREME, A. Quantitative trait dissection analysis in Eucalyptus using RAPD markers: I. Detection of QTL in Interespecific hybrid progeny, stability of QTL, expression across different ages. . **Theoretical and Applied Genetics**. v. 95, p. 597-608, 1997.
- VIJAYALAKSHMI S., YADAV K., KUSHWAHA C., SARODE S.B., SRIVASTAVA C.P., CHAND R., SINGH B.D. Identification of RAPD markers linked to the rust (*Uromyces fabae*) vresistance gene in pea (*Pisum sativum*). **Euphytica**, v.144, p.265-274, 2005.
- WEIGUO L., MING Y.Z., ENRIQUE A.P., CALVIN F.K. Highly efficient doubled-haploid production in wheat (*Triticum aestivum* L.) via induced microspore embryogenesis. **Crop Science**, v.42, p.682-692, 2002.
- WEIR B.S., **Genetic data analysis II**. Sinauer Associates, Sunderland, 1996, 445p.
- WILCOX P.L., AMERSON H.V., KUHLMAN E.G., LIU B., O'MALLEY D.M., SEDEROFF R.R. Detection of a major gene for resistance to fusiform rust disease in loblolly pine by genomic mapping. **Proceedings of the National Academy of Sciences**, v.93, p.3859-3864, 1996.
- YIN T. M., DIFAZIO S.P., GUNTER L.E., RIEMENSCHNEIDER D., TUSKAN G.A. Large-scale heterospecific segregation distortion in *Populus* revealed by a dense genetic map. **Theoretical and Applied Genetics**. v. 109, p. 451-463, 2001.
- YOUNG A., BOYLE T.J.B., BOSCHER D. **Forest conservation genetics: Principles and practice**. Australia: CSIRO, 2000. 352p.
- YU K., PARK S. J., POYSA V. Marker assisted selection of common beans for resistance to common bacterial blight: efficacy and economics. **Plant breeding**, v.199, p.411-415, 2000.
- ZOBEL, B., JETT J. B. **Genetics of wood production**. Berlin: Springer-Verlag, 1995. 337p.

Capítulo I

Proposta de metodologia para mapeamento de locos controladores de características oligogênicas - Populações de retrocruzamento

1. RESUMO

Características oligogênicas se caracterizam por sua herança governada por genes de maior efeito e por sua importância para as espécies vegetais cultivadas, com destaque para a resposta de resistência das plantas às doenças. A natureza discreta e a interação epistática destas características resultam em um padrão de herança que não deve ser interpretado pelos métodos tradicionais de detecção de QTL. A busca por métodos que permitam melhor interpretar a informação genética produzida durante a recombinação continua sendo um dos principais objetivos do mapeamento e identificação de QTLs. O objetivo deste trabalho é propor um método para mapeamento e detecção de locos controladores da expressão de características oligogênicas, OTLs (“Oligogenic Trait Loci”) em populações de retrocruzamento. Este método, definido como M.M.C.O (Método de Mapeamento de Características Oligogênicas), utiliza funções de verossimilhança para obtenção de estimativas de recombinação “ r ” de melhor ajuste aos dados, a partir das probabilidades condicionais de ocorrência entre o loco marcador e a herança da característica oligogênica. Os resultados mostram que o método foi adequado para detecção de OTLs em populações de retrocruzamento. Os OTLs foram corretamente caracterizados e posicionados no mapa em populações de tamanho limitado, compostas por até 200 indivíduos. Análises comparativas entre as metodologias mostram que as estimativas M.M.C.O. de “ r ” entre o loco marcador e os locos OTLs são mais acuradas do que as obtidas pelos métodos de marca e de intervalo simples. O M.M.C.O. não necessita de informação prévia de ordenamento dos marcadores e não é influenciado pela ocorrência de OTLs em um mesmo grupo de ligação. O atendimento das pressuposições para análise das características oligogênicas, a maior acurácia das estimativas obtidas a partir da informação contida em todas as classes genotípicas e a possibilidade de posicionar pontualmente o OTL fazem do M.M.C.O. uma estratégia de potencial para contribuir na interpretação da herança das características oligogênicas.

2. ABSTRACT

Oligogenic traits are distinguished by their heritage ruled by higher effect genes and by their importance for the cultivated plants, with emphasis to the plant disease resistance heritage. The qualitative nature and epistatic interaction of these traits results in a heritage pattern that should not be interpreted through traditional methods for QTL (Quantitative Trait Loci) detection. The investigation for methods which allow a better interpretation of the genetic information produced during the recombination is still one of the major objectives of QTL detection. The objective of this work is to propose a method for detection of important loci related with the expression of oligogenic traits – OTL (Oligogenic Trait Loci). This method, defined as Oligogenic Trait Mapping Method (O.T.M.M.), uses maximum likelihood probability functions to obtain the r estimates of better adjustment to the observed data from the conditional probabilities of occurrence between the molecular markers and the OTL. The results show that the method was adequate for OTL detection in backcrossing populations. Comparative analysis with other QTL detection methodologies shows that O.T.M.M. has the more accurate r estimates of the distance between the molecular markers and the OTL loci. The O.T.M.M. does not need previous information of the markers order in the linkage groups and are not influenced by the presence of two QTL in the same linkage group. The focus in the analysis presumes the oligogenic heritage, the higher accuracy of the r estimates, obtained from the information contained in all genotypic classes, and the position of the OTL in the map, contribute to make O.T.M.M. a powerful tool for the analysis of oligogenic traits.

3. INTRODUÇÃO

O desenvolvimento do ramo das ciências biológicas que se destina ao estudo da herança e variação dos caracteres tem permitido, associado às técnicas modernas de estudo da molécula do DNA, o entendimento dos mecanismos hereditários com uma riqueza de detalhes nunca antes experimentada (DEKKERS & HOSPITAL, 2001). O mapeamento genético utilizando técnicas de marcadores moleculares permite a decomposição da herança complexa em fatores mendelianos não influenciados pelo ambiente (MURO ABAD et al. 2005; SCHUSTER & CRUZ et al. 2004; ROCHA et al. 2003; MACKAY, 2001). O avanço de novas técnicas de análise genômica têm proporcionado um contínuo acúmulo de informações genômicas em novas espécies de plantas e animais (GRATTAPAGLIA, 2004).

O monitoramento de regiões do genoma visando prever padrões de expressão herdáveis baseia-se na pressuposição de que os locos contribuem diferencialmente para a resposta das características (GELDERMANN, 1975). A identificação destas regiões genômicas de potencial para auxiliar procedimentos de seleção é um dos principais desafios para a prática da seleção assistida (VARSHNEY et al. 2005; DIAS et al. 2005; BADU et al. 2004; YOUNG, 1999). Vários trabalhos têm demonstrado que o aumento na precisão dos procedimentos de seleção proporcionado pela incorporação da informação genômica têm acelerado e, em alguns casos, reduzido os custos de obtenção de cultivares e espécies melhoradas (ALZATE-MARIN et al. 2005; BOUCHEZ et al. 2002; ROCHA et al. 2002; YOUSEF & JUVIK, 2001; MOREAU, et al. 2000; YU et al. 2000; LANDE & THOPSON; 1990).

O sucesso de experimentos que visam à identificação de genes candidatos depende do atendimento dos requisitos necessários para assegurar a qualidade experimental necessária ao fornecimento de estimativas confiáveis de ligação fatorial (YOUNG, 1999). Vários são os fatores determinantes da qualidade dos mapas de ligação, entre os quais se destacam: a população avaliada (ROCHA et al. 2004, LI, 2004; DARVASSI & SOLLER, 1995, GRATTAPAGLIA & SEDEROFF; 1994), o tamanho da população (SILVA; 2005), a técnica de marcador molecular empregada (MATTHEWS, et al. 2001; LIU, 1997; WEIR, 1996), a precisão experimental (SCHUSTER & CRUZ; 2004,

CRUZ et al. 2003; JAUHAR, 2000) e a herdabilidade da característica avaliada (CRUZ & SILVA, 2006). Além destes, a natureza da herança da característica, se monogênica, oligogênica ou quantitativa, é determinante para a definição da estratégia experimental mais adequada (SCHUSTER & CRUZ, 2004, LIU, 1997, LANDER & BOTSTEIN, 1989).

Características oligogênicas, de distribuição discreta e expressão governada por poucos genes de maior efeito, têm se mostrado importantes na condução dos programas de melhoramento de várias culturas (RAO & LI, 2000; ZOBEL & JETT, 1995). Para o melhoramento vegetal, a resistência das plantas à ação de patógenos é provavelmente a característica mais importante de natureza oligogênica (ALZATE-MARIN, et al. 2005; ALFENAS et al. 2004). A resistência ao *Ceratocystis* (ALFENAS et al. 2004) e a resistência à Ferrugem (VIJAYALAKSHMI et al. 2005; ALFENAS et al. 2004; JUNGHANS et al. 2003; WILCOX et al. 1996) são exemplos de algumas características de interesse para o melhoramento vegetal e de provável herança oligogênica.

Vários métodos para a identificação de QTLs de características de distribuição contínua estão bem caracterizados, no entanto, menos alternativas estão disponíveis para a análise de locos controladores de características oligogênicas - OTL ("Oligogenic Trait Loci"). A natureza discreta da resposta associada à interação epistática de múltiplos fatores resulta em uma característica que não deveria ser mapeada pelos métodos tradicionais, que pressupõem normalidade e herança governada por muitos genes de pequeno efeito (LI, 2004; ZENG, 1994; JANSEN, 1993; LANDER & BOTSTEIN, 1989). Esta limitação já foi abordada na literatura em outros trabalhos. VISSCHER et al. (1996) compararam métodos utilizando regressão linear e regressão linear generalizada a partir de dados binários e concluíram que não havia diferença na acurácia da predição e no poder do teste das metodologias. XU e ATCHLEY (1996) estenderam a metodologia de mapeamento por intervalo composto para dados binários utilizando modelos de regressão logística. RAO e LI (2000) apresentaram estratégias para análise de variáveis discretas utilizando diferentes modelos alélicos. E ainda existem autores que apresentam aproximações Bayesianas para detecção do OTL (YI et al. 2004). Contudo, nenhuma das estratégias pesquisadas aborda a questão da maneira considerada neste trabalho.

O objetivo deste trabalho é apresentar um método adicional para detecção de locos controladores da expressão de características controladas por genes de maior efeito (OTL), com ênfase em segregações e interações epistáticas governadas por dois e três genes em populações de retrocruzamento. O Método de Mapeamento de Características Oligogênicas (M.M.C.O) se baseia na utilização de funções de verossimilhança para prever novas possibilidades de segregação entre marcadores moleculares e locos controladores de características oligogênicas visando à obtenção de estimativas de distância entre os marcadores e o loco OTL.

4. MATERIAL E MÉTODOS

4.1 Simulação de dados

Foi utilizado o módulo de simulação do aplicativo computacional GQMOL (CRUZ, 2007), para a obtenção de populações de retrocruzamento de uma espécie diplóide com 4 grupos de ligação de 200 cM de tamanho cada um, saturação média de 5 cM com marcadores de natureza co-dominante.

4.1.1 Simulação do genoma

Visando avaliar os indivíduos recombinantes provenientes de uma configuração genômica mais parcimoniosa para as análises de recuperação da informação biológica paramétrica, foi considerado um genoma constituído de quatro grupos de ligação com 30 marcas eqüidistantes co-dominantes e nível de saturação médio de 5 cM (Figura 1). Os grupos de ligação G.L.1, G.L.2 e G.L.3 possuem locos que alteram a expressão da característica (marcador C15, G.L.1, marcador C215 G.L.2, marcador C326 G.L. 3). O OTL presente no grupo de ligação 3 só influencia a expressão da característica quando esta apresenta interações epistáticas entre três genes. O grupo de ligação 4 não apresenta em nenhuma das situações consideradas locos que alteram a expressão da característica.

4.1.2 Simulação dos genitores

Todas as populações simuladas foram obtidas a partir de eventos de recombinação provenientes do mesmo arranjo de genes na população F1, tendo a segregação de genitores completamente contrastantes resultado somente na produção de eventos de recombinação em fase de acoplamento.

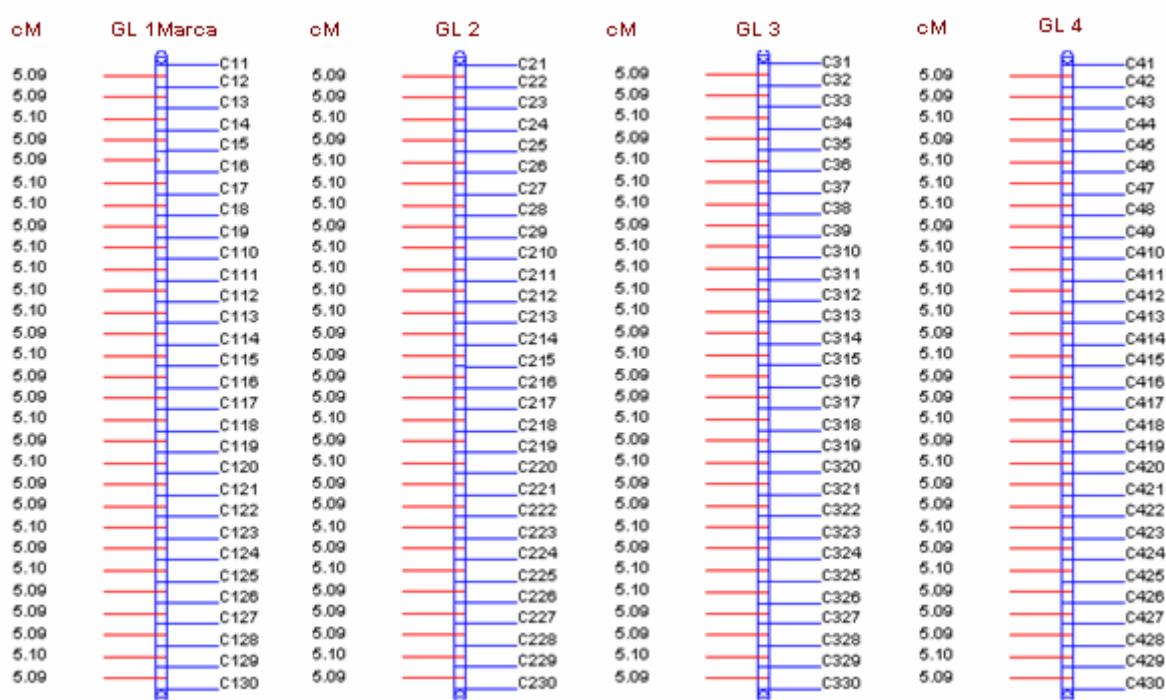


Figura 1 – Genoma com 4 grupos de ligação (G.L.) e saturação média de 5 cM, contendo 30 marcadores por grupo de ligação utilizado para obtenção dos genótipos dos genitores e das populações segregantes.

4.1.3 Simulação da característica oligogênica

Para caracterizar a metodologia proposta foram simuladas duas características governadas por dois e três genes de interação epistática com proporções esperadas entre as classes (1) e (0) de 3:1 e 7:1 (Tabela 1). Para simulação da epistasia entre os genes foram considerados dois marcadores do genoma original controladores da característica oligogênica C1 (marcador C15, G.L.1, marcador C215 G.L.2) e três marcadores controladores da característica oligogênica C2 (marcador C15, G.L.1, marcador C215 G.L.2, marcador C326 G.L. 3). A Tabela 2 mostra presença e ausência dos marcadores e a resposta nas variáveis C1 e C2. Para as análises de comparação entre os métodos foram geradas (100 populações) constituídas por 50, 100 e 200 indivíduos.

Tabela 1 – Genótipos e proporções esperadas considerando epistasia do tipo dominante dos alelos A, B e C dos genes controladores das características oligogênicas avaliadas.

Epistasia dominante 2 genes		Epistasia dominante 3 genes	
3/4	AaBb	7/8	AaBbCc
	Aabb		AaBbcc
	aaBb		AabbCc
1/4	aabb		aaBbCc
			aabbCc
			aaBbcc
			Aabbcc
		1/8	aabbcc

Tabela 2- Relação entre a presença dos marcadores e as respostas nas variáveis C1 (3:1) e C2 (7:1). O fenótipo da característica C1 tem expressão influenciada pela presença dos marcadores C15-G.L.1, C215-G.L.2 e o da C2 têm expressão influenciada pelos marcadores C15-G.L.1, C215-G.L.2, C326-G.L. 3.

G.L.	Marcador								
1	5	1	1	1	0	1	0	0	0
2	15	1	1	0	1	0	1	0	0
3	26	1	0	1	1	0	0	1	0
Fenótipo C1		R	R	R	R	R	R	S	S
Fenótipo C2		R	R	R	R	R	R	R	S

1,0 – Representa a presença ou ausência do marcador molecular, 1-presença, 0 ausência; R,S- Representa a expressão da característica oligogênica R – resistente, S – Suscetível.

4.2 Análises genômicas

4.2.1 Mapeamento genético

No intuito de avaliar o método proposto para recuperação da informação de herança de características oligogênicas, as análises descritas a seguir foram consideradas.

4.2.2 Análise de segregação de locos individuais – teste de (χ^2)

Foram aplicados testes de qui-quadrado (χ^2) para verificar a razão de segregação de cada marca em todas as populações geradas. A estatística qui-quadrado é dada por:

$$(i) \quad \chi^2_{calculado} = \sum_i \left[\frac{(O_i - E_i)^2}{E_i} \right]$$

4.2.3. Análise de pares de marcas - estimação da porcentagem de recombinação

As estimativas de recombinação foram obtidas pelo método analítico definido por SCHUSTER & CRUZ (2004) que consiste na utilização da função de verossimilhança para a estimação dos valores de distância do mapa.

$$(ii) \quad L(p_i / n_i) = \lambda p_1^{n_1} p_2^{n_2} p_3^{n_3} \dots p_n^{n_n}$$

Em que :

$$(iii) \quad \lambda = \frac{N!}{n_1! n_2! \dots n_n!}$$

$$(iv) \quad N = \sum_{i=1}^n n_i$$

Em que:

N é o tamanho da população

n_i é o número de ocorrências do evento x_i com probabilidade p_i .

4.2.4 - Análise de pares de marcas - determinação dos grupos de ligação e ordenamento das marcas

A determinação dos grupos de ligação foi obtida pelo agrupamento baseado na frequência máxima de recombinação ($r_{\max}$) e o LOD mínimo ($\text{LOD}_{\min}$). Neste trabalho foram considerados os valores de $r_{\max} = 30\%$ e $\text{LOD}_{\min} = 3$. Após a formação de um grupo de ligação preliminar verificou-se a melhor ordem dos n marcadores constituintes do grupo pelo método SARF ("Sum of Adjacent Recombination Fractions").

4.3. Comparação de genomas

4.3.1. Detecção de características oligogênicas

Funções de verossimilhança foram utilizadas para obtenção de estimativas de distância entre os marcadores e os locos controladores da expressão destas características. O teste estatístico para testar a ligação entre o marcador e a característica oligogênica é o da razão de verossimilhança que consiste na comparação entre o estimador de máxima verossimilhança θ_A e a probabilidade de ocorrência de θ_N sob a hipótese nula de que $\theta_N = 0.5$; sendo 0.5 a distância entre o marcador e loco controlador da característica oligogênica.

$$(v) \quad L \frac{(\theta_A / x)}{(\theta_N / x)}$$

Para facilidade de interpretação, muitos autores preferem utilizar o logaritmo na base 10 da razão de verossimilhança:

$$(vi) \quad Z = \text{Log}_{10} \left[L \frac{(\theta_A / x)}{(\theta_N / x)} \right]$$

O valor do estimador de Z significa que a hipótese alternativa é 10^Z vez mais provável que a hipótese nula. A consistência, normalidade, eficiência assintótica e as estimativas de variância, que podem ser estimadas pelo inverso do índice de informatividade, são propriedades desejadas e que caracterizam os estimadores de máxima verossimilhança.

5 - RESULTADOS E DISCUSSÃO

5.1- Simulação dos dados

O programa de análises genômicas GQMOL (CRUZ, 2007) permite gerar populações segregantes a partir da informação do genoma e dos genótipos de genitores. A sequência do processo de simulação é dividida nas seguintes etapas: i) determinação do tamanho e número dos grupos de ligação, saturação média do mapa e número de marcadores; ii) geração de indivíduos parentais contrastantes para os marcadores; iii) produção dos gametas (conjunto de 10.000 gametas recombinantes obtidos considerando a possibilidade de permuta intervalo a intervalo); e iv) combinação ao acaso dos gametas para obtenção dos genótipos da população segregante.

O processo aleatório de ocorrência de permuta intervalo a intervalo possui a probabilidade de ocorrência igual à distância entre as marcas, visando simular o pareamento dos cromossomos homólogos e a ocorrência de permuta ao longo do grupo de ligação nos intervalos delimitados por duas marcas adjacentes. Este processo sequencial é contínuo até que todos os intervalos entre marcadores adjacentes dos grupos de ligação tenham contribuído para a formação do gameta recombinante (CRUZ, 2007; SILVA, 2005).

A informação genotípica individual é obtida a partir do genótipo dos pais e do genoma referência criada para análise. Neste trabalho, a escolha da ploidia referência deve-se ao fato de as espécies diplóides serem bastante representativas das espécies vegetais cultivadas (BOREM, 2001). A natureza co-dominante do marcador foi escolhida considerando a natureza das técnicas mais recentes utilizadas para mapeamento (BUNYAMIN et al. 2002, BRONDANI et al. 1998, CHARLESWORTH et al. 1994). Quatro grupos de ligação são suficientes para explorar os fenômenos de epistasia que serão estudados e o valor de saturação escolhido é um valor conservador observado nos principais trabalhos de mapeamento. Considera-se um mapa com este valor de saturação como medianamente saturado (MACKAY, 2001). Muitos trabalhos sugerem que o número de 200 indivíduos deve ser considerado mínimo para o desenvolvimento de mapas de ligação (SILVA, 2005, SCHUSTER & CRUZ, 2004, MACKAY, 2001).

Para avaliar a utilidade da metodologia M.M.C.O., foram consideradas populações provenientes de retrocruzamento para mapeamento da característica oligogênica. Além dos aspectos práticos relacionados ao cruzamento controlado e à multiplicação dos indivíduos, as populações de retrocruzamento destacam-se pela rapidez de obtenção em campo e a contribuição genética de apenas um genitor na segregação das progênes, fazendo com que o estudo baseado nestas populações seja útil para o mapeamento e a detecção de QTLs em diversas culturas (ROCHA et al. 2004; SCHUSTER & CRUZ, 2004).

Com este objetivo, foi simulada uma população de retrocruzamento composta por 200 irmãos completos provenientes do cruzamento entre a F1 e o genitor portador de alelos recessivos, saturação média de 5 cM e duas características oligogênicas de expressão binária e natureza epistática, C1 e C2, governadas respectivamente pelas marcas C15 (G.L.1) e C215 (G.L.2) e pelas marcas C15 (G.L.1), C215 (G.L.2) e C326 (G.L.3). As condições deste estudo visam obter conclusões que sejam aplicadas, tendo em vista que muitas pesquisas de mapeamento realizadas com 200 indivíduos e saturação média de 5 cM almejam o desenvolvimento de mapas-referência (MISSIAGIA, 2005, NOVAES, 2005, MEKSEM et al. 2001, CERVERA et al. 2001, SILVA, 2005, LIU, 1997).

5.2- Estimativas de probabilidade condicional

A utilização de funções de verossimilhança permite a obtenção de estimativas de variância mínima estimada a partir da informação contida em todas as classes genóticas que compõem a população, (SCHUSTER & CRUZ, 2004, LYNCH & WALSH, 1998). Os estimadores de probabilidade condicional expressam a razão da ocorrência conjunta das classes genóticas do OTL e do marcador e são necessários para a estimação dos valores de recombinação utilizando o método de máxima verossimilhança (LIU, 1997). As estimativas de probabilidade apresentadas a seguir consideram as pressuposições em relação ao padrão de herança: i) expressão binária do tipo presença ou ausência e ii) herança governada por poucos genes e segregação correlacionada com uma marca molecular previamente mapeada. Os valores

de probabilidade de ocorrência conjunta da marca molecular e do fenótipo da característica oligogênica foram representados em função das classes fenotípicas de resistência (R) e suscetibilidade (r) associadas à presença de uma marca molecular (M) ou à sua ausência (m) (Tabelas 3 e 4).

A seguir são apresentadas as estimativas de probabilidade de ocorrência das classes genotípicas em relação ao loco marcador (M) e os locos controladores das características (A e B, ou A, B e C) considerando a herança governada por 2 ou por 3 genes e os tipos de interação epistáticas mais comuns em populações de retrocruzamento 3:1 (Tabela 3) e 7:1 (Tabela 4). Os estimadores de probabilidade da ocorrência individual dos locos permitem a obtenção das estimativas de probabilidades de segregação condicional entre o marcador e os locos controladores da resistência.

Tabela 3 - Proporção teórica esperada das classes genotípicas em população de retrocruzamento com padrões de segregação esperado de 3:1 da característica oligogênica.

Classes genotípicas		Probabilidade em função de r	Classes Fenotípicas 3:1
MmAa	Bb	$(1-r)/4$	1R
	bb	$(1-r)/4$	1R
Mmaa	Bb	$r/4$	1R
	bb	$r/4$	1S
mmAa	Bb	$r/4$	0R
	bb	$r/4$	0R
mmaa	Bb	$(1-r)/4$	0R
	bb	$(1-r)/4$	0S

M,m- Formas alternativas do marcador molecular M; A,a- Formas alternativas do gene A; B,b- Formas alternativas do gene B; 0,1- Presença ou ausência da marca molecular; R,S- Fenótipo da característica oligogênica – Resistente ou suscetível.

O somatório das proporções esperadas de ocorrência da resposta da característica oligogênica (3:1) e a presença do marcador molecular são apresentados nas expressões a seguir:

$$(vii) \quad P(1R) = \frac{2-r}{4}$$

$$(viii) \quad P(1S) = \frac{r}{4}$$

$$(ix) \quad P(0R) = \frac{r+1}{4}$$

$$(x) \quad P(0S) = \frac{1-r}{4}$$

Tabela 4 - Proporção teórica esperada das classes genotípicas em população de retrocruzamento com padrões de segregação esperado de 7:1 da característica oligogênica.

Classes genotípicas			Freqüências esperadas em função de r	Classes Fenotípicas 7:1
MmAa	Bb	Cc	$(1-r)/8$	1R
	Bb	cc	$(1-r)/8$	1R
	bb	Cc	$(1-r)/8$	1R
	bb	cc	$(1-r)/8$	1R
mmaa	Bb	Cc	$r/8$	1R
	Bb	cc	$r/8$	1R
	bb	Cc	$r/8$	1R
	bb	cc	$r/8$	1S
mmAa	Bb	Cc	$r/8$	0R
	Bb	cc	$r/8$	0R
	bb	Cc	$r/8$	0R
	bb	cc	$r/8$	0R
mmaa	Bb	Cc	$(1-r)/8$	0R
	Bb	cc	$(1-r)/8$	0R
	bb	Cc	$(1-r)/8$	0R
	bb	cc	$(1-r)/8$	0S

M,m- Formas alternativas do marcador molecular M; A,a- Formas alternativas do gene A; B,b- Formas alternativas do gene B; C,c- Formas alternativas do gene C; 0,1- Presença ou ausência da marca molecular; R,S- Fenótipo da característica oligogênica – Resistente ou suscetível.

O somatório das proporções esperadas de ocorrência da resposta da característica oligogênica (7:1) e a presença do marcador molecular são mostrados nas expressões a seguir:

$$(xi) \quad P(1R) = \frac{4-r}{8}$$

$$\begin{aligned} \text{(xii)} \quad & P(1S) = \frac{r}{8} \\ \text{(xiii)} \quad & P(0R) = \frac{3+r}{8} \\ \text{(xiv)} \quad & P(0S) = \frac{1-r}{8} \end{aligned}$$

A inclusão de outros valores de probabilidade condicional associados a padrões de segregação específicos permite adaptar as funções de verossimilhança a quaisquer outros tipos de populações e características. Segundo SCHUSTER & CRUZ (2004), as funções de verossimilhança são de grande importância para o desenvolvimento de mapas de ligação e para a detecção de QTLs, uma vez que poucas pressuposições devem ser consideradas. A utilização da função de verossimilhança permite a obtenção de estimativas de variância mínima, estimadas a partir da informação contida em todas as classes genotípicas que compõem a população, propriedades estas de grande relevância, principalmente quando comparada a métodos mais simples de estimação, como a contagem da proporção dos indivíduos recombinantes em relação ao total. O maior número de cálculos necessários para a obtenção destas funções é a principal limitação para o uso desta ferramenta (WEIR, 1996; LIU, 1997).

5.3- Funções de verossimilhança

Além das estimativas de probabilidade condicional, é necessário para estimação do valor de recombinação o conhecimento da função densidade de probabilidade (f.d.p.) da distribuição da variável aleatória (SCHUSTER & CRUZ, 2004, LIU, 1997; WEIR, 1996) . Desta maneira, considera-se que a probabilidade de ocorrência das classes genotípicas nas populações segregantes assume distribuição multinomial (LIU, 1997). As funções de verossimilhança utilizadas para estimar o valor de r entre marcadores e QTL para as duas características oligogênicas consideradas são apresentadas abaixo:

Característica oligogênica (3:1)

$$(xxx) L(pi / ni) = \lambda \left(\frac{2-r}{4} \right)^{n_1} \left(\frac{r}{4} \right)^{n_2} \left(\frac{1+r}{4} \right)^{n_3} \left(\frac{1-r}{4} \right)^{n_4}$$

Em que :

$$\lambda = \frac{N!}{n_1!n_2!\dots n_n!}$$

$$N = \sum_{i=1}^n n_i$$

(xxxi) Característica oligogênica (7:1)

$$L(pi / ni) = \lambda \left(\frac{4-r}{8} \right)^{n_1} \left(\frac{r}{8} \right)^{n_2} \left(\frac{3+r}{8} \right)^{n_3} \left(\frac{1-r}{8} \right)^{n_4}$$

Em que :

$$\lambda = \frac{N!}{n_1!n_2!\dots n_n!}$$

$$N = \sum_{i=1}^n n_i$$

5.4. Mapeamento genético

Segundo LIU (1997), a obtenção de um mapa genético de qualidade é pré-requisito para a detecção de QTLs. Neste contexto, a interpretação dos dados de mapeamento foi realizada visando inferir a qualidade do mapa estimado a partir dos parâmetros definidos para a simulação.

Previamente à análise dos pares de marcas, os marcadores tiveram sua segregação individual testada pelo teste de qui-quadrado com correção de Bonferroni para testes múltiplos (dados não mostrados). Nenhuma das marcas avaliadas apresentou segregação distorcida segundo o teste de qui-quadrado. Considerando a análise total de 120 marcadores, é esperado que em média ($\alpha < 0,01$) 1,2 marcas apresentem distorção ao acaso na avaliação de n populações. O estudo por simulação permite o isolamento do efeito do acaso como principal fonte de interferência na produção dos gametas e na recombinação dos genótipos, considerando-se, desta maneira, que mesmo se algum marcador tivesse apresentado segregação distorcida, ele não deveria ser descartado (LIU, 1997; WEIR, 1996).

Para a obtenção dos grupos de ligação, foram utilizados os parâmetros de “LOD score” mínimo (LOD = 3,0) e máxima frequência de recombinação ($r = 0,30$), em função dos valores tradicionalmente considerados nos trabalhos de mapeamento. Observa-se que o valor mais elevado do “LOD score” (LOD = 30) poderia ser considerado sem resultar em uma redução no número de marcas que compõem os grupos de ligação (LOD_{mínimo} = 34,95 observado entre as marcas C311 e C312). A interpretação do valor de LOD pelo teste da razão de verossimilhança indica que, para esta população, a hipótese de ligação entre os marcadores adjacentes em um mesmo grupo de ligação é no mínimo 10^{30} vezes mais provável que a hipótese de segregação independente entre estas marcas (LANDER & BOTSTEIN, 1989). Este valor indica forte evidência de ligação fatorial entre as marcas, uma vez que é qualitativamente superior ao nível mínimo sugerido (LOD=3) para rejeição da hipótese de segregação independente.

Outros trabalhos sugerem a suficiência dos parâmetros considerados na simulação para a obtenção das populações de retrocruzamento. SCHUSTER & CRUZ (2004) avaliaram a acurácia das estimativas de recombinação obtidas

em populações de retrocruzamento compostas por 200 indivíduos. A avaliação do conteúdo médio de informação e dos intervalos de confiança associados às estimativas de recombinação indica que a população de retrocruzamento é uma das que conferem as estimativas de recombinação mais precisas juntamente com as populações de F2 em aproximação mapeada com marcadores co-dominantes e população de duplos haplóides. (ROCHA et al. 2004)

Após a obtenção das estimativas de recombinação entre marcadores adjacentes, a ordem mais provável dos marcadores foi determinada utilizando a metodologia “SARF” (*“sum of adjacent recominant fractions”*). Foram recuperados 4 grupos de ligação originais com saturação média de 5.12 cM (Figura 1). A Tabela 6 mostra a média dos valores de distâncias e os valores de tamanho dos grupos de ligação comparados com os valores paramétricos, sendo os valores de desvios de acordo com os dados obtidos por (SILVA, 2005) para 200 indivíduos. Não houveram marcas não ligadas. Os resultados indicam que os parâmetros definidos para simulação das populações, o número de indivíduos e a saturação média do mapa, determinados considerando a natureza prática dos trabalhos de mapeamento, permitiram a obtenção de estimativas de recombinação de robustez suficiente para recuperação da informação paramétrica.

Tabela 6- Comparação entre tamanhos estimados dos grupos de ligação e valores de distância média entre as marcas estimadas a partir da informação de 200 genótipos recombinantes de população de retrocruzamento.

Grupo de ligação	Tamanho estimado	Tamanho paramétrico	Desvio	Média estimada	Média paramétrica	Desvio
1	154,27	150	4,27	5,319655	5	0,319655
2	151,14	150	1,14	5,211724	5	0,211724
3	143,58	150	-6,42	4,951034	5	-0,04897
4	143,58	150	-6,42	4,951034	5	-0,04897

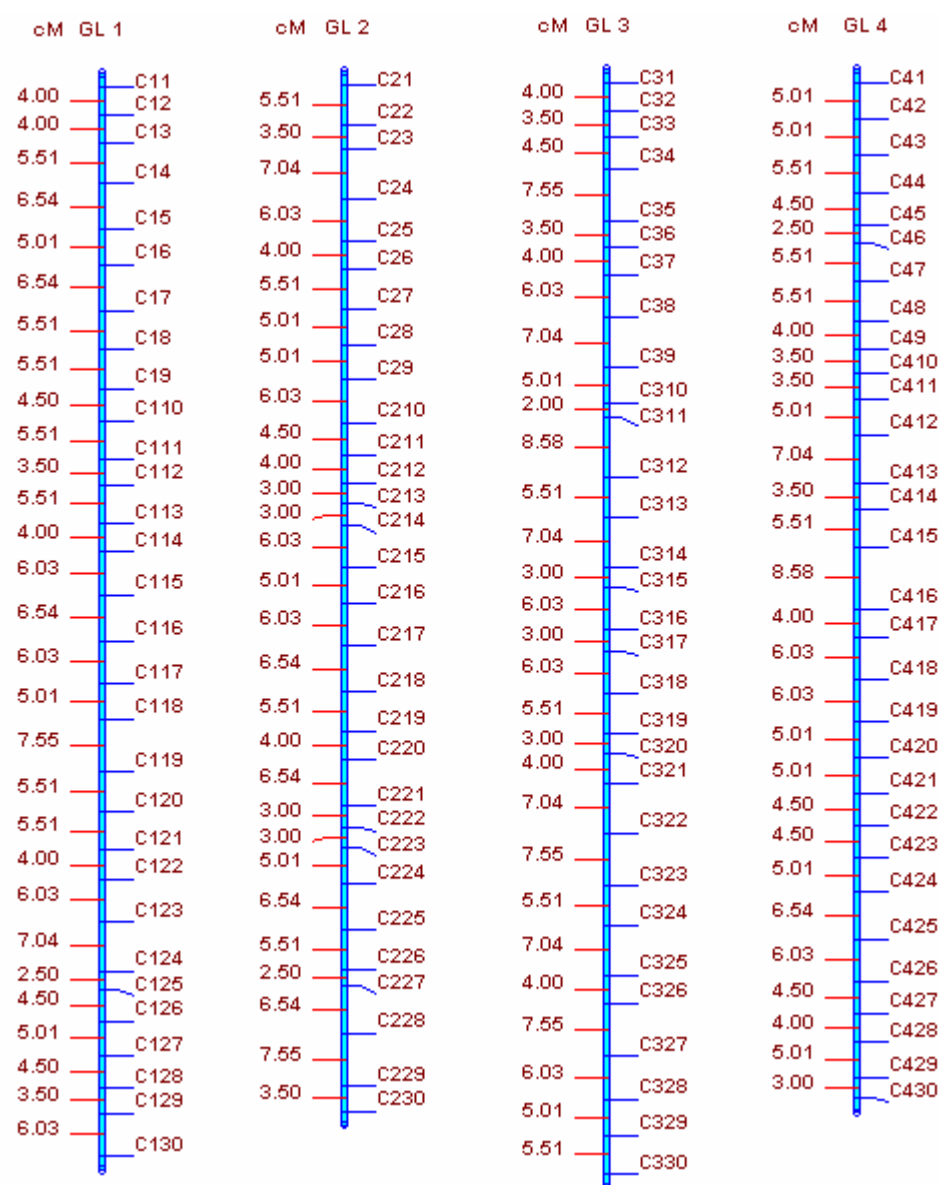


Figura 2 - Mapa genético de população de retrocruzamento, estimada a partir da informação de recombinação entre os marcadores provenientes de 200 genótipos obtidos a partir de combinação aleatória de 10.000 gametas recombinantes, produzidos por simulação, utilizando o aplicativo computacional GQMOL. Os grupos de ligação foram obtidos utilizando-se valores de LOD = 3 e $r = 0,30$ e unidade de Kosambi.

5.5. Detecção do OTL

A definição dos marcadores de maior importância para a expressão da característica baseia-se em dois critérios: valor de r mínimo entre os locos dos marcadores moleculares e do OTL ($r_{\text{mínimo}} = 10$) e valor de “LOD score” que testa a hipótese de ligação contra a hipótese de segregação independente. A Tabela 7 mostra os valores de distância estimados entre os marcadores e o loco OTL e seus respectivos valores da estatística “LOD score”. Utilizando esta tabela, é possível detectar as marcas moleculares de maior importância para a expressão das características oligogênicas. A observação dos dados mostra oito marcadores de efeitos significativos para a expressão do C1 de expressão governada por 2 locos e 11 marcadores de efeito significativo para a expressão do C2 de expressão governada por 3 locos. Uma simples observação na Tabela 7 mostra a correta detecção dos locos controladores das características consideradas (C1- marcador C15, G.L.1, marcador C215 G.L.2, C2 - marcador C15, G.L.1, marcador C215 G.L.2, marcador C326 G.L. 3).

O conhecimento da posição dos marcadores mapeados nos grupos de ligação é pré-requisito dos métodos de mapeamento por intervalo. A não necessidade do conhecimento deste ordenamento prévio para utilização do método M.M.C.O. é bastante atrativa, uma vez que muitos estudos de mapeamento demonstram a ocorrência de marcadores não ligados para a expressão das características (ROCHA et al. 2007).

Após a identificação dos putativos grupos aos quais devem pertencer os locos OTL, estes grupos têm sua ordem novamente questionada pelo método SARF, incluída a informação da característica oligogênica considerada como uma marca adicional no mapa, visando obter a melhor estimativa de posição do OTL ao longo dos grupos de ligação. O método SARF tem se mostrado particularmente confiável mesmo na ocorrência de desbalanceamentos em função de dados perdidos de genotipagem (SCHUSTER & CRUZ, 2004, LIU, 1997; WEIR, 1996). Na ocorrência de empate entre os ordenamentos, como observado no grupo de ligação 1, a melhor ordem é eleita ao acaso entre aquelas de menor somatório de valores de distância entre marcadores adjacentes.

Uma vez eleita à ordem de maior parcimônia entre os marcadores e os

locos controladores da característica oligogênica, o loco OTL pode ser posicionado no mapa como mostrado na Figura 2. A Figura 2 mostra o mapa genético original em azul e em verde os grupos de ligação acrescidos da posição dos OTLs. Observa-se que a incorporação do loco OTL ao mapa de ligação deve resultar em modificações nos valores de recombinação diretamente proporcionais à magnitude dos valores da distância entre marcas adjacentes e inversamente proporcionais ao número de eventos meióticos analisados.

Acredita-se que a consideração das pressuposições em relação ao número de locos e da natureza da interação entre alelos e entre genes resulte em um aumento na precisão e poder de teste da metodologia em relação àquelas tradicionalmente utilizadas para detecção de QTLs. Além do método proposto, a população foi avaliada quanto à detecção do QTL utilizando outros métodos de detecção de QTLs: marca simples (M.S.) e intervalo simples (I.S.). A subjetividade relacionada com a escolha dos co-fatores do método de mapeamento por intervalo composto fez esta metodologia inapropriada nas comparações seguintes.

A Figura 3 mostra os resultados das técnicas de detecção de QTL da marca simples e o intervalo simples utilizado para a detecção de QTL, mostrando também que todas as metodologias foram capazes de detectar corretamente os marcadores candidatos em população de retrocruzamento composta por 200 indivíduos. No entanto, considerando que as condições experimentais são as principais limitações dos estudos de mapeamento genético, a comparação entre metodologias de análises é também objeto de vários estudos de mapeamento genético que objetivam interpretar com a máxima acurácia e eficiência a informação genética contida nos padrões de segregação dos delineamentos genéticos.

Tabela 7: Estimativas de recombinação obtidas pelo método M.M.O para as características C1 e C2 associadas aos marcadores posicionados nos grupos de ligação. Os marcadores candidatos estão realçados.

Marca	GL	Dist(cM)	LOD	Dist(cM)	LOD	Marca	GL	Dist(cM)	LOD	Dist(cM)	LOD
		C 1		C 2				C 1		C 2	
C11	1	13,9	7,31	16	2,98	C31	3	48,2	0,02	48,3	0,01
C12	1	11,6	8,36	11,9	3,86	C32	3	48,2	0,02	48,3	0,01
C13	1	8,8	10,02	11,5	4,05	C33	3	48,8	0,01	44,4	0,07
C14	1	6,5	11,61	7,4	5,24	C34	3	46	0,08	41,8	0,16
C15	1	0	16,98	0	8,45	C35	3	46,5	0,06	35,6	0,47
C16	1	5	11,77	4,1	6,02	C36	3	46	0,08	32,6	0,68
C17	1	10,1	8,82	8,4	4,61	C37	3	48,2	0,02	35,1	0,51
C18	1	16,9	5,95	16,6	2,81	C38	3	48,2	0,01	37,4	0,37
C19	1	16,5	6,15	16,3	2,89	C39	3	49,4	0	42,8	0,12
C110	1	18,4	5,47	20	2,24	C310	3	49,4	0	38,4	0,31
C111	1	25,1	3,19	21,9	1,88	C311	3	48,2	0,01	37,4	0,37
C112	1	24,6	3,34	21,5	1,95	C312	3	44,2	0,17	29,2	1,01
C113	1	24,3	3,49	21,1	2,02	C313	3	51,6	0,01	34,1	0,59
C114	1	27,8	2,58	28,8	1,06	C314	3	52,5	0,07	42,8	0,12
C115	1	25,6	3,16	24,6	1,56	C315	3	52,5	0,1	46,1	0,03
C116	1	26,4	2,98	19,4	2,39	C316	3	52,5	0,13	38,4	0,31
C117	1	27,3	2,71	24,1	1,62	C317	3	52,5	0,07	34,1	0,59
C118	1	30,8	1,94	27,5	1,23	C318	3	52,5	0,19	40,5	0,2
C119	1	33,6	1,41	32,1	0,75	C319	3	52,5	0,17	44,4	0,07
C120	1	38,6	0,67	36,9	0,4	C320	3	52,5	0,2	45,5	0,04
C121	1	35,6	1,05	34,1	0,59	C321	3	52,5	0,23	44,4	0,07
C122	1	35,6	1,05	34,1	0,59	C322	3	52,5	0,17	40,1	0,22
C123	1	35,8	1,04	32,1	0,75	C323	3	52,5	0,04	28,8	1,06
C124	1	34,6	1,22	28,8	1,06	C324	3	45,4	0,11	14	3,12
C125	1	34,1	1,31	28,4	1,12	C325	3	46,5	0,06	9	4,3
C126	1	38,1	0,74	36,4	0,43	C326	3	43,1	0,24	0	7,7
C127	1	39,3	0,6	39,6	0,25	C327	3	37,9	0,74	4,6	5,45
C128	1	36,4	0,96	34,9	0,54	C328	3	41,3	0,38	14,3	3,03
C129	1	34,7	1,21	31,3	0,84	C329	3	43,6	0,2	21,9	1,88
C130	1	32,5	1,6	23	1,82	C330	3	47,6	0,03	25,8	1,37
C21	2	27,1	2,64	23,5	1,61	C41	4	47,1	0,04	42,8	0,12
C22	2	25,6	3,05	22,2	1,81	C42	4	48,2	0,02	41,5	0,16
C23	2	28,8	2,33	25,4	1,44	C43	4	52,5	0,04	50,6	0
C24	2	27,8	2,58	24,6	1,56	C44	4	48,8	0,01	44,4	0,07
C25	2	25,8	2,94	19,8	2,1	C45	4	49,4	0	44,9	0,06
C26	2	24,8	3,23	18,9	2,25	C46	4	51,1	0,01	46,6	0,03
C27	2	21,7	4,06	14	3,12	C47	4	48,2	0,02	43,8	0,08
C28	2	21,7	4,06	14	3,12	C48	4	47,6	0,03	43,2	0,1
C29	2	16,8	5,67	9,4	4,1	C49	4	45,4	0,11	45,5	0,04
C210	2	19,1	5,1	12,4	3,67	C410	4	42,5	0,28	42,8	0,12
C211	2	16,5	6,15	12,1	3,77	C411	4	40,8	0,43	38,9	0,28
C212	2	15,7	6,55	11,6	3,96	C412	4	40,8	0,43	43,3	0,1
C213	2	13,5	7,52	11,6	3,96	C413	4	42	0,33	42,2	0,14
C214	2	9,6	9,29	8,1	4,82	C414	4	44,8	0,14	42,8	0,12
C215	2	0	15,47	0	7,82	C415	4	47,6	0,03	47,7	0,01
C216	2	6,4	9,91	9,6	4	C416	4	51,6	0,01	49,4	0
C217	2	9,3	8,59	14,3	3,03	C417	4	47,1	0,04	40,6	0,2
C218	2	11,6	7,65	18,6	2,33	C418	4	47,1	0,04	40,6	0,2
C219	2	17,3	5,47	28,1	1,09	C419	4	43,6	0,2	35,1	0,51
C220	2	18,9	5,01	27,1	1,2	C420	4	44,8	0,14	40,5	0,2
C221	2	16,9	5,95	25	1,5	C421	4	50,6	0	50,6	0
C222	2	20	4,75	25,8	1,37	C422	4	45,4	0,11	45,5	0,04
C223	2	23,7	3,64	29,2	1,01	C423	4	43,6	0,2	46,1	0,03
C224	2	27,3	2,71	24,1	1,62	C424	4	45,9	0,08	46,1	0,03
C225	2	27,8	2,58	28,8	1,06	C425	4	48,8	0,01	46,6	0,03
C226	2	28,3	2,45	29,2	1,01	C426	4	52,5	0,03	50	0
C227	2	28,8	2,33	34,1	0,59	C427	4	48,2	0,02	43,7	0,08
C228	2	29,3	2,21	34,6	0,55	C428	4	45,9	0,08	46,1	0,03
C229	2	30,9	1,86	31,6	0,77	C429	4	44,7	0,14	40,4	0,2
C230	2	28	2,48	22,2	1,81	C430	4	45,9	0,08	39,2	0,25

G.L. – Grupo de Ligação; Dist (cM)- Distância entre o marcador e o OTL; LOD – Valor da estatística LOD das estimativas de r entre marcador – OTL.

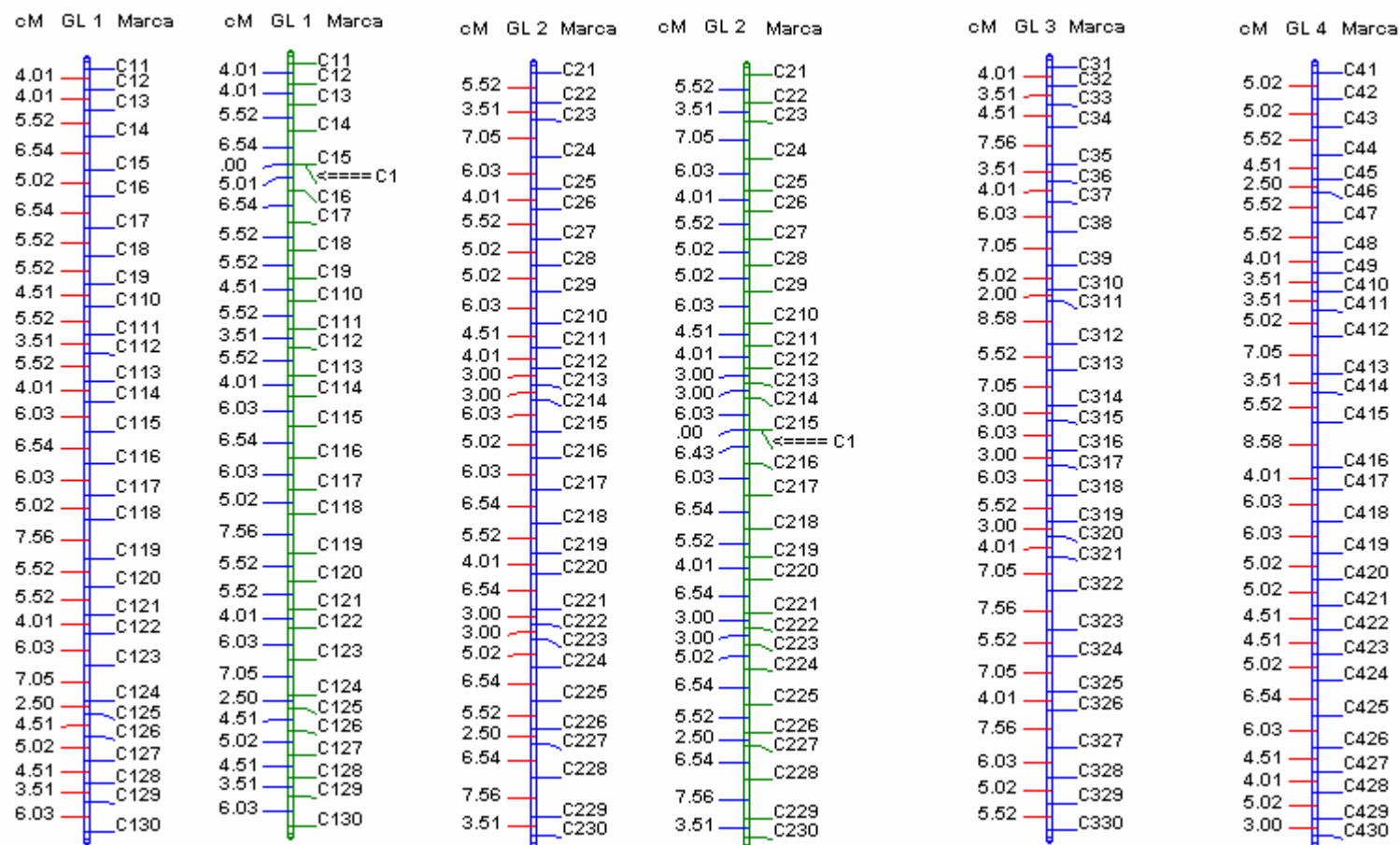


Figura 3 - Mapa genético de população de retrocruzamento, estimado a partir da informação de recombinação entre os marcadores provenientes de 200 genótipos. Os grupos de ligação ($LOD = 3, r = 0,30$ e unidade de Kosambi) foram obtidos pelo método analítico de mapeamento e aos grupos de ligação realçados em verde foi acrescida a informação de distância e posição dos locos controladores da expressão da característica oligogênica

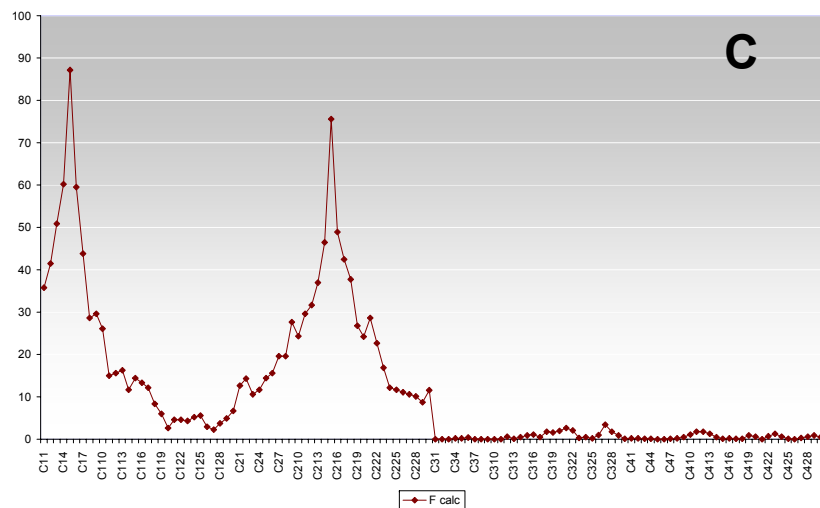
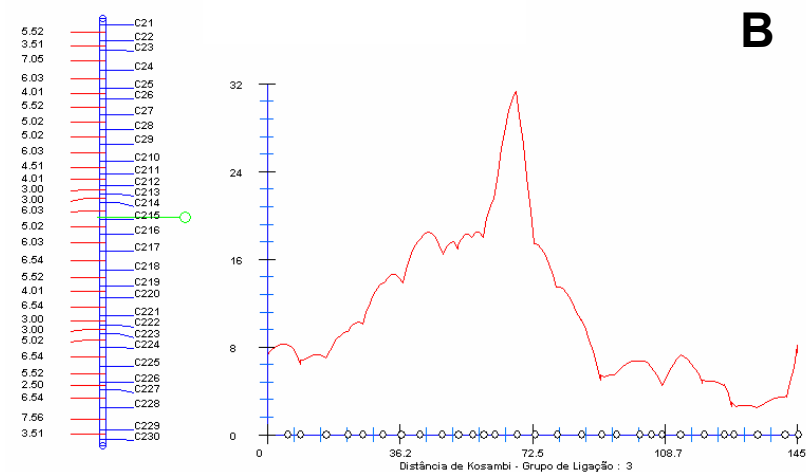
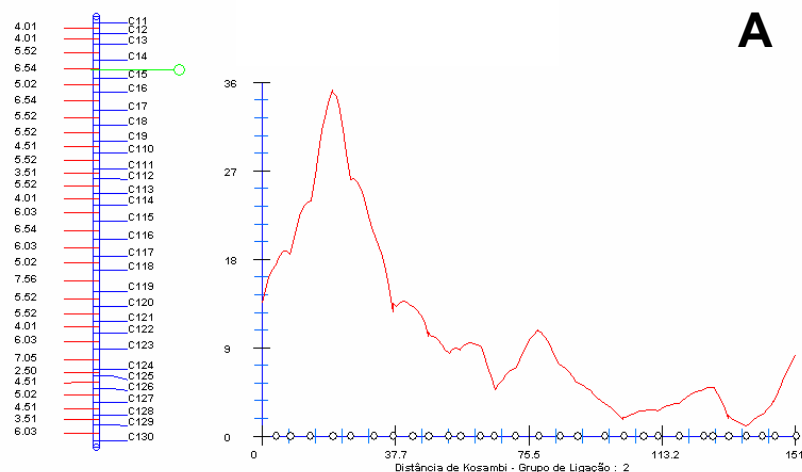


Figura 4 – Resultados das técnicas de mapeamento por intervalo simples (A e B) e distribuição dos valores de F do método da marca simples (C) da avaliação da característica C1, em população de retrocruzamento composta de 200 indivíduos.

5.6. Comparação entre métodos: marca simples (M.S.) e mapeamento por intervalo simples (I.S.)

Se o número de indivíduos e outras condições experimentais não fossem limitantes para a pesquisa, métodos menos robustos poderiam ser utilizados para a detecção de locos controladores das características. A busca por novos métodos e estratégias de análise é assunto atual e pertinente para o mapeamento genético de plantas (YOUSEF e JUVIK, 2001, DEKKERS & HOSPITAL, 2001; LANDE & THOMPSON, 1990).

Entre os métodos de propriedades mais bem caracterizadas à disposição do pesquisador, destacam-se do mais simples para o mais complexo: mapeamento por marcas simples, mapeamento por intervalo simples e o mapeamento por intervalo múltiplo (SCHUSTER & CRUZ, 2004). Visando à comparação com o método proposto, foram simuladas 100 populações de retrocruzamento compostas por 50, 100 e 200 indivíduos e uma única característica governada por três genes, sendo que dois dos locos que governam a expressão destas características estão alocados no mesmo grupo de ligação (marcadores C15 e C120, grupo de ligação 1) e um terceiro loco (marcador C215, grupo de ligação 2).

5.6.1. Critérios para comparação entre os métodos

Dois tipos de erros estão associados ao teste de hipótese. O erro tipo I que pode ser definido como a não rejeição da hipótese H_0 quando esta é verdadeira. O erro tipo II definido como a aceitação da hipótese nula quando esta é falsa (STEEL & TORRIE, 1980). Neste contexto, o poder do teste é definido como a probabilidade de rejeitar a hipótese nula quando a hipótese alternativa é verdadeira (STEEL & TORRIE, 1980). A avaliação dos marcadores candidatos permite a comparação entre os métodos, uma vez que o número de QTLs não detectados pode ser interpretado como uma estimativa da ocorrência de erro tipo I e a identificação de QTLs em posições equivocadas do genoma uma estimativa da

ocorrência de erro tipo II.

O M.M.C.O. apresenta algumas propriedades semelhantes ao método de marcas simples: i) baseia-se na análise de apenas uma marca por vez e ii) não necessita da informação prévia de ordenamento das marcas dentro dos grupos de ligação. Três critérios foram considerados para comparação entre estas duas metodologias: 1- o número de marcadores candidatos indicados pelo teste F (marca simples) e pela estatística “LOD score” (funções de verossimilhança). 2 – a distância entre os locos candidatos e o QTL, se r maior ou menor que 50 unidades de mapa 3 - o número de vezes que estas metodologias não foram capazes de detectar os locos controladores da expressão e 4- a acurácia das estimativas de recombinação. Neste contexto, considera-se “marcador candidato” o conjunto de tamanho variável de marcadores de segregação correlacionada com o loco QTL de acordo com as metodologias testadas (MACKAY, 2001).

O método de mapeamento por intervalo simples, de natureza diferenciada em relação à estratégia apresentada, baseia-se em um modelo de regressão múltipla, de resposta gráfica, preferido por muitos autores para interpretação. Para comparação com o método de mapeamento por intervalo simples, dois critérios foram considerados: 1- o número de vezes que os locos controladores das características se encontravam em pontos de máximo da função e 2- o número de falsos QTLs encontrados.

5.6.2. Marca simples e M.M.C.O.

A Figura 5 mostra os valores de média de 100 populações constituídas por 50, 100 e 200 indivíduos avaliados em relação ao número de marcadores candidatos detectados pela marca simples (M.S.) e pelo método de mapeamento de características oligogênicas (M.M.C.O.). O M.M.C.O. mostra tendência de aumento do número de marcadores candidatos ($r \leq 50$) e de manutenção no número de marcadores candidatos ($r \geq 50$) com o aumento no tamanho da população. Esta tendência é resultado do aumento na acurácia do método em função da análise maior número de eventos meióticos.

Considera-se que marcadores candidatos com distância superior a 50 unidades de mapa representam a rigor uma associação espúria entre os locos marcadores e o loco OTL. Comparativamente, a Tabela 9 mostra que as funções de verossimilhança indicam menos marcadores candidatos com $r \geq 50$ em relação à marca simples. Como observado na Figura 5, também não há alteração nas proporções dos marcadores significativos a 5, pelo teste da marca simples com o aumento da população.

Tabela 9 - Número de marcadores candidatos detectados pela M.S. e M.M.C.O.

Método		Tamanho da população		
		50	100	200
MS(5%)	$r \geq 50$	0,65	0,72	0,75
	$r \leq 50$	0,35	0,28	0,25
MS(1%)	$r \geq 50$	0,66	0,59	0,26
	$r \leq 50$	0,34	0,41	0,74
M.M.C.O.(LOD≥ 3)	$r \geq 50$	0,11	0,23	0,09
	$r \leq 50$	0,89	0,77	0,91

O método da marca simples possui limitações bem caracterizadas, como o confundimento entre o efeito do QTL e sua distância (SCHUSTER & CRUZ, 2004). A avaliação das populações indica que aproximadamente 30% dos marcadores candidatos significativos a 5% de probabilidade não apresentam ligação fatorial com o loco QTL e não deveriam ser considerados para subsidiar nenhuma decisão de pesquisa futura, independentemente do tamanho da população. Os resultados também indicam de maneira consistente que o número de marcadores significativos a 1% e 5% detectados pelo método da marca simples aproxima-se quatro vezes do valor de marcadores candidatos detectados pelo M.M.C.O.

Comparativamente à marca simples, as funções de verossimilhança apresentaram maior número de locos OTL não detectados, aproximando-se este número de zero com o aumento da população (Tabela 8). Os valores dos desvios destas variáveis também indicam um aumento na acurácia do método em relação ao aumento do número de indivíduos analisados (Figura 5).

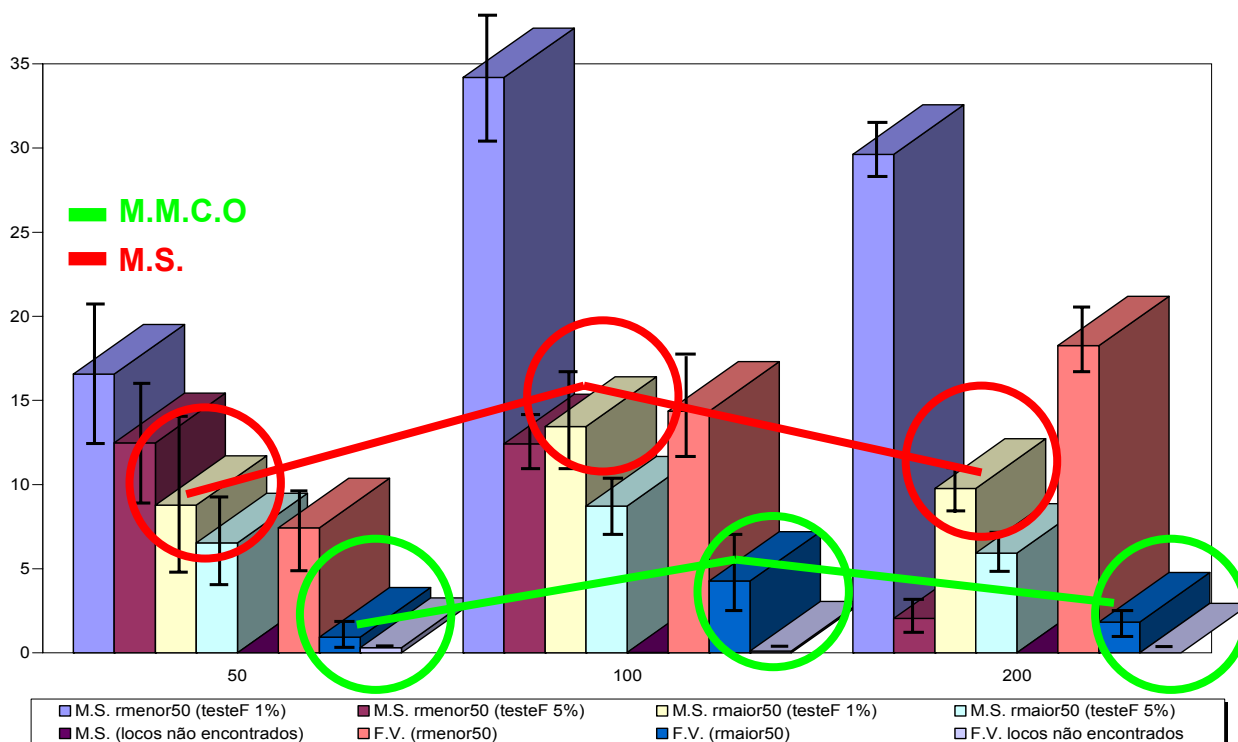


Figura 5 – Média de 100 simulações do número de marcadores candidatos indicados pelos M.M.C.O. e de marca simples em populações de retrocruzamento construídas de 50, 100 e 200 indivíduos. O número de marcadores candidatos com $r \geq 50$ está realçado. A barra transversal indica o valor do desvio relacionado com as observações.

Tabela 8 – Locos não encontrados pelas metodologias de marca simples (M.S.) e pelo método de mapeamento de características oligogênicas (M.M.C.O.) avaliados em 100 populações de retrocruzamento.

OTLs não encontrados	Tamanho da população		
	50	100	200
M.S.	0	0	0
M.M.C.O.	0,16	0,05	0

** Os valores representam o percentual de locos não encontrados em relação ao número total de locos (200).

Em comparação com o método M.S., o M.M.C.O. indica um conjunto mais restrito de marcadores candidatos. Este grupo mais restrito de marcas está em geral contido no conjunto de marcadores candidatos eleitos pelo método da marca

simples, uma vez que a distribuição dos valores da estatística F da marca simples e dos valores de LOD das funções de verossimilhança apresenta expressivo valor de correlação ($r \cong 1$). A interpretação dos resultados das funções de verossimilhança em conjunto com a marca simples permite filtrar de um conjunto inicial de marcas aquelas de maior potencial de segregação correlacionada ($r \leq 50$) com o loco OTL.

Além da indicação dos marcadores candidatos, provavelmente, um dos critérios mais importantes de comparação entre estas duas metodologias seja a acurácia das estimativas de recombinação. Por permitir a obtenção de estimativas de recombinação entre o loco marcador e o OTL, de melhor ajuste a partir da informação contida em todas as classes genotípicas, é de se esperar que as estimativas M.M.C.O. sejam mais acuradas do que as estimativas provenientes da M.S. que testam apenas o contraste entre duas classes de marcadores. Desta maneira, em populações de retrocruzamento, os testes de hipóteses do M.M.C.O. estão associados a 3 graus de liberdade, em comparação com a M.S., associada a apenas 1 grau de liberdade.

A Figura 6 mostra a dispersão das estimativas M.M.C.O. e M.S. de r em comparação com os valores paramétricos de r representados pela linha azul. Por definição, quanto menor o valor de r mais próximo o loco marcador está do loco OTL. A dispersão dos valores das estimativas de r mostra um melhor ajuste das estimativas M.M.C.O. dos valores paramétricos.

Para intervalos de 10 u.m. (unidades de mapa) na vizinhança do loco OTL, as estimativas M.M.C.O. apresentam pequenos desvios em relação aos valores paramétricos e de maneira bastante acurada, identificaram os dois locos de maior importância relacionados com a expressão da característica C1 e os três locos de maior importância relacionados com a característica C2.

O maior poder de teste, acurácia das estimativas associado à não necessidade de conhecimento prévio de ordenamento entre as marcas, faz o método M.M.C.O mais indicado para análise de características oligogênicas do que o método da M.S.

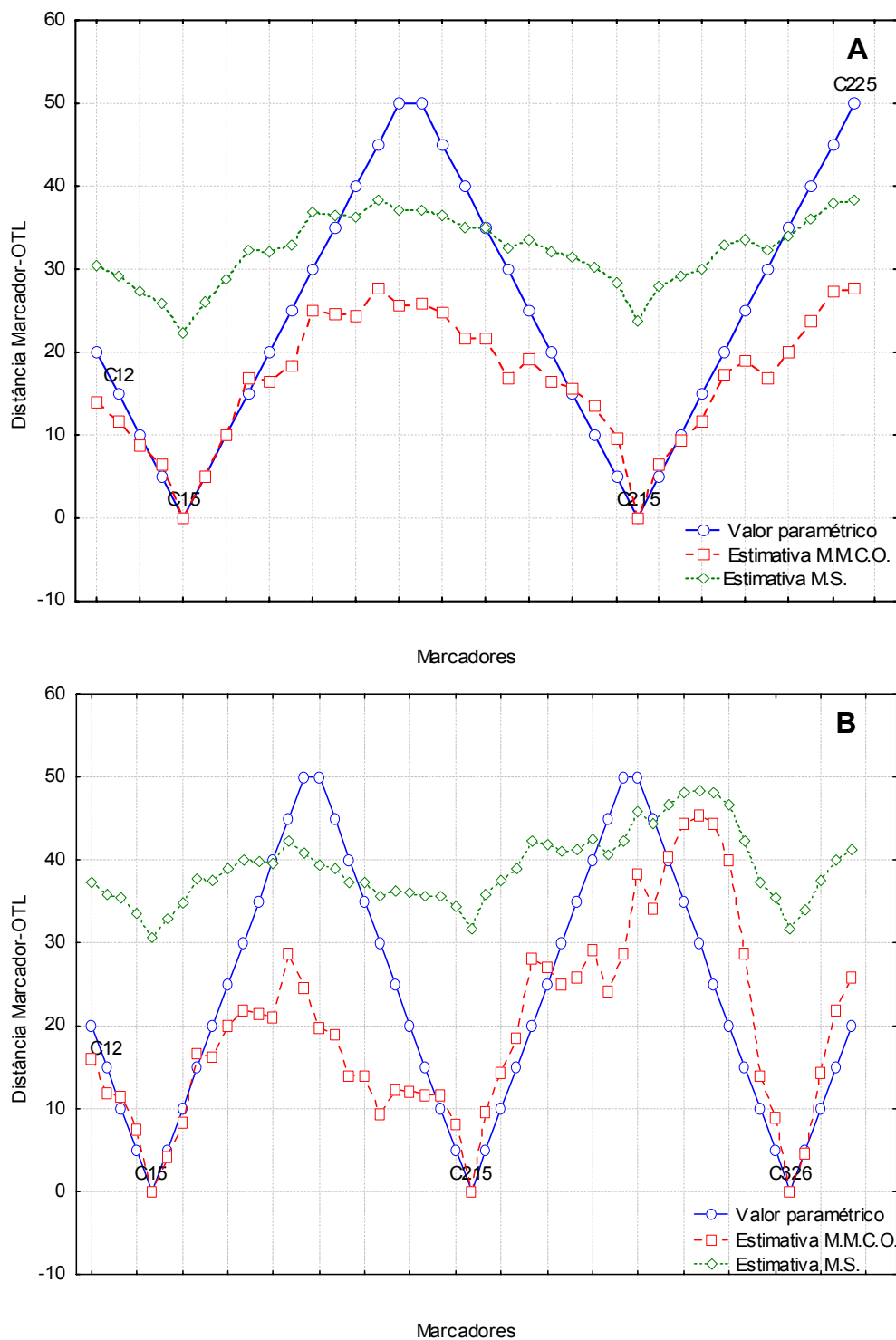


Figura 6 – Dispersão das estimativas de r obtidas pelo método M.M.C.O. e da M.S.

A linha azul representa os valores paramétricos de distância entre as marcas moleculares e o loco OTL. 6A mostra a dispersão dos valores relativos à característica C1 de expressão controlada por dois locos OTL e 6B mostra a dispersão relativa à característica C2 de expressão controlada por 3 locos OTL.

5.6.3. Mapeamento por intervalo simples (I.S.) e M.M.C.O.

A comparação entre M.M.C.O. e I.S. foi realizada considerando apenas o número de falsos OTLs encontrados, tendo em vista a dificuldade de comparação entre metodologias de naturezas diferentes. Observou-se a ocorrência de QTLs fantasmas em todas as repetições analisadas pelo método do intervalo simples, de maneira consistente, independentemente do número de indivíduos considerados. Além desta limitação já bem caracterizada do método (ZENG, 1994), a presença de regiões com valor de LOD superior a 3 com distância superior a 50 unidades de mapa do OTL também foi observada em aproximadamente 80% das repetições avaliadas. A Figura 6 ilustra uma das avaliações de uma das populações de retrocruzamento constituída de 200 indivíduos analisada neste trabalho.

O M.M.C.O. possui propriedades específicas que podem ser utilizadas em conjunto de maneira sinérgica com outras metodologias de análise. A maior acurácia e precisão das estimativas associadas individualmente ao marcador permitem uma avaliação mais criteriosa dos resultados provenientes das análises de mapeamento por intervalo.

6. CONCLUSÕES

Ainda que na literatura estejam disponíveis conjuntos mais amplos de métodos do que os avaliados neste trabalho, não existe um único método que se aplique bem à dados de qualquer natureza. A utilização criteriosa das ferramentas mais apropriadas permite ao pesquisador certificar-se de suas conclusões e avaliar a confiabilidade de seus resultados face aos dois tipos de erro associados ao teste de hipótese. Conclusões equivocadas, relacionadas à ocorrência do erro tipo II, são as que potencialmente podem causar mais danos à aplicação dos resultados de mapeamento. A utilização das funções de verossimilhança para recuperação da informação genômica paramétrica indica o caráter mais restritivo deste método que pode ser agregado a outras metodologias para diminuir a

incidência de conclusões equivocadas em relação ao posicionamento do loco OTL.

Se as condições experimentais não forem limitantes, a comparação das metodologias mostrará que as estratégias tradicionalmente utilizadas para mapeamento podem ser utilizadas para detectar características governadas por poucos genes. O M.M.C.O. apresenta propriedades complementares a outras metodologias e pode ser considerado como um método adicional para a caracterização de regiões mais importantes do genoma. As funções de verossimilhança utilizadas no M.M.C.O. são mais restritivas do que a marca simples, mais acuradas, não necessitando de ordenamento prévio das marcas e não apresentando a limitação de confundimento entre OTL s adjacentes em um mesmo grupo de ligação. A facilidade relativa de discriminação das proporções epistáticas, resultado da ação de dois ou três genes em populações de retrocruzamento, faz desta metodologia uma alternativa bastante atrativa para a caracterização da herança de características oligogênicas, principalmente considerando que a utilização dos métodos tradicionais não permite a caracterização dos padrões de herança mais simples e mendelianos como observados na herança das características oligogênicas.

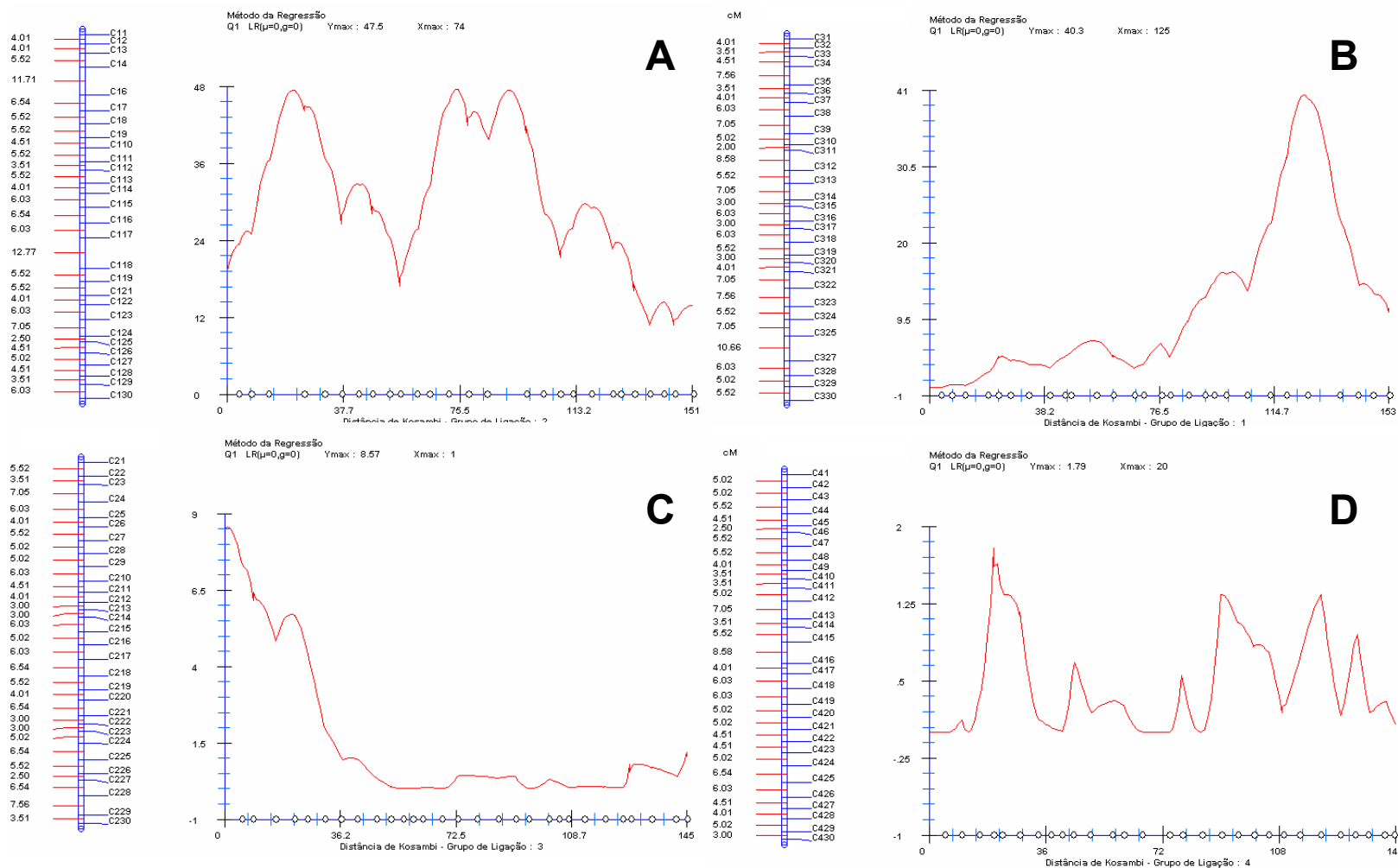


Figura 6 – Exemplo de avaliação de uma das 100 repetições avaliadas mostrando a ocorrência de QTL fantasma no grupo de ligação 1 em uma região com LOD 9 no grupo de ligação 3 que não possui nenhum loco de expressão correlacionada com a característica.

7. REFERÊNCIAS BIBLIOGRÁFICAS

- ALFENAS A.C., ZAUZA, E.A.V., MAFIA R.G., ASSIS T.F. **Clonagem e doenças do Eucalipto**. Viçosa, MG. Editora UFV, 2004. 442p.
- ALZATE-MARIN A.L., CERVIGNI G.D.L., MOREIRA M.A., BARROS E.G. Seleção assistida por marcadores moleculares visando ao desenvolvimento de plantas resistentes a doenças, com ênfase em feijoeiro e soja. **Fitopatologia Brasileira**, v. 30(4), p. 333-342, 2005.
- BADU R., SUDHA K.N., PRASANNA B.M., GUPTA H.S. Integrating marker assisted selection in crop breeding – Prospects and challenges. **Current Science**, v.87, n.5, p.607-619, 2004.
- BOREM, A. **Melhoramento de plantas**. Editora UFV. 3ª Edição. Viçosa, MG. UFV, 2001. 500 p.
- BOUCHEZ A., HOSPITAL F., CAUSSE M., GALLAIS A., CHARCOSSET A. Marker-assisted introgression of favorable alleles at quantitative trait loci between maize elite lines. **Genetics** v.162: p.1945-1959, 2002.
- BRONDANI, R.P.V., BRONDANI C., TARCHINI R., GRATTAPAGLIA D. Development, characterization and mapping of microsatellite markers in *Eucalyptus grandis* and *E. urophylla* . **Theoretical and Applied Genetics**, v.97, p.816-827, 1998.
- BUNYAMIN T., THOMAS E.M., PAULS K.P. Genetic mapping of agronomic traits in common bean. **Crop Science**, v.42, p.544-556, 2002.
- CERVERA M.T., STORME V., IVENS B., GUSMÃO J., LIU B.H., HOSTYN V., SLYCKEN J.V., MONTAGU M.V., BOERJAN W. Dense Genetic Linkage Maps of Three *Populus* Species (*Populus deltoides*, *P. nigra* and *P. trichocarpa*) Based on AFLP and Microsatellite Markers. **Genetics**, v.158 p.787-809, 2001.
- CHARLESWORTH, B., SNIEGOWSKI, P., STEPHAN, W. The evolutionary dynamics of repetitive DNA in eukaryotes. **Nature**, v.371, p.215 – 220, 1994.
- CRUZ, C. D. **Programa para análises de dados moleculares e quantitativos – GQMOL**. Viçosa: UFV, 2007.(Software em desenvolvimento).
- CRUZ C.D., REGAZZI A.J., CARNEIRO P.C.S. **Modelos Biométricos Aplicados ao Melhoramento Genético - Volume 1**, Viçosa , MG. Editora UFV, 2004. 480p.

- CRUZ C.D., REGAZZI A.J., CARNEIRO P.C.S. **Modelos Biométricos Aplicados ao Melhoramento Genético - Volume 2**, Viçosa , MG. Editora UFV, 2003. 585p.
- CRUZ C.D., SILVA L. C. **Análise de marcadores moleculares**. In: BORÉM A., Marcadores moleculares, Viçosa , MG. Editora UFV, 2006. 374p
- DARVASI A, SOLLER M. Advanced intercross lines, an experimental population for fine genetic mapping. **Genetics**, v.141, p.1199-1207, 1995.
- DEKKERS J.C.M., HOSPITAL F. The use of molecular genetics in the improvement of agricultural populations. **Nature reviews**, v.3, p.22-32, 2001.
- DIAS, L.A.S., ROCHA, R. B., PICOLI, E.A.T Distinctness of cacao cultivars using yield component data and RAPD markers. **Crop breeding and Applied Biotechnology**, v.5, p.47 - 54, 2005.
- GELDERMANN, H. Investigations on inheritance of quantitative characters in animals by genes markers. **Theoretical and Applied Genetics**., v. 46:p.310-330, 1975.
- GRATTAPAGLIA, D. Integration genomics into the *Eucalyptus* breeding. **Genetics and Molecular Research**. v. 3 (3), p. 369-379, 2004.
- GRATTAPAGLIA, D., SEDEROFF, R. Genetic linkage maps of *Eucalyptus grandis* and *E. urophylla* using a pseudo-testcross mapping strategy and RAPD markers. **Genetics**, v.137, p.1121-1137, 1994.
- JANSEN, R.C. Interval mapping of multiple quantitative trait loci. **Genetics**. v.135, p.205 -211, 1993.
- JAUHAR P. P., **Methods of genomes analysis in plants**. Florida: C.R.C. Press, Inc., 2000. 386p.
- JUNGHANS D. T., ALFENAS A. C., BROMMONSHENKEL S. H., ODA S., MELLO E. J., GRATTAPAGLIA D. Resistance to rust (*Puccinia psidii* Winter) in Eucalyptus mode of inheritance and mapping of a major gene with RAPD markers. **Theoretical and Applied Genetics**. v.108, 175-180. 2003.
- LANDE R., THOMPSON R. Efficiency of Marker-Assisted Selection in the Improvement of Quantitative Traits. **Genetics**, v.124, p.743-756, 1990.
- LANDER E. S., BOTSTEIN D., Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. **Genetics** v.121, n.1, p185-199, 1989

- LI J., **Development of multiple interval mapping for mapping QTL in ordinal traits**. Raleigh, North Carolina, North Carolina State University, 2004. 124p.
- LIU B.H., **Statistical Genomics: linkage mapping and QTL analysis**. Boca raton, Florida, USA: CRC Press, 1997. 610p.
- LYNCH, M.; WALSH, B. **Genetics and analysis of quantitative traits**. Sunderland, USA: Sinauer Associates, Inc.Ed., 1998., 978p.
- MACKAY T.F.C., Genetic architecture of quantitative traits. **Annual Review of Genetics**. v.35, p.303-339; 2001.
- MATTHEWS F.B., DEVINE T.E., WEISEMANN J.M., BEARD H.S., LEWERS K.S., MACDONALD M.H., PARK Y., MAITI R., LIN J., KUO J., PEDRONI M.J., CREGAN P.B. , SAUNDERS J.A. Incorporation of sequenced cDNA and genomic markers into the soybean genetic map. **Crop Science**, v.41, p.516-521, 2001.
- MEKSEM, K., NJITI, V.N., BANZ, W.J., IQBAL, M.J., KASSEM, M.M., HYTEN, D.L.,YUANG, J., WINTERS, T.A., LIGHTFOOT, D.A. Genomic regions that underlie soybean seed isoflavone content. **Journal of Biomedicine and Biotechnology**, v1, n.1, p. 38-44, 2001.
- MISSIAGIA A. A., **Mapeamento Genético de QTL para qualidade da madeira e florescimento precoce e estudos de expressão gênica alelo específica em *Eucalyptus* spp.** Piracicaba, 2005. 236p. Dissertação (Doutorado), Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo.
- MOREAU, L., LEMARIE, S., CHARCOSSET, A., GALLAIS, A. Economic efficiency on one cycle of marker-assisted selection. **Crop Science**, v.40, p.329-337, 2000.
- MURO ABAD, J.I., ROCHA, R.B., CRUZ, C.D., ARAÚJO, E. F. Crosses recommendation method for obtaining *Eucalyptus* spp. hybrids assisted by molecular markers. **Crop Breeding Applied Biotechnology**, v.5, p.71 - 73, 2005.
- NOVAES E. **Mapeamento de QTLs para qualidade da madeira em *Eucalyptus grandis* x *E. urophylla* e ancoragem de clones BAC no mapa genético.** Viçosa, 2006. 171p. Dissertação (Mestrado), Universidade Federal de Viçosa.
- RAO, S., LI X., Strategies for genetic mapping of quantitative traits. **Genetica**, v.109, p.183–197, 2000.

- ROCHA R.B., MURO ABAD, J.I., ELEOTÉRIO I.P., ARAÚJO E.F. Fingerprint and genetic diversity analysis of *Eucalyptus spp.* genotypes using RAPD and SSR markers, **Scientia Forestalis** v.62, p.24-31, 2002.
- ROCHA R.B., PEREIRA J.F., CRUZ C.D., QUEIROZ M.V., ARAÚJO E.F. O mapeamento genético no melhoramento de plantas. **Biotecnologia Ciência e Desenvolvimento**. v. 30, p.27-32, 2003.
- ROCHA, R.B., CRUZ, C.D., BARROS, W. S., FERREIRA, F.M., ARAÚJO, E.F. Comparisons of segregating populations for genetic mapping. **Crop Breeding and Applied Biotechnology**., v.4, p.408 - 415, 2004.
- ROCHA, R. B., CRUZ, Cosme Damião, ROSADO, A. M., BARROS, E. G., ARAÚJO, Elza Fernandes de Mapping of QTLs related with wood quality and development characteristics in hybrids (*Eucalyptus grandis* x *Eucalyptus urophylla*). (Artigo no prelo). **Revista Árvore**, 2007.
- SILVA, L. C. **Simulação do tamanho da população e da saturação do genoma para mapeamento genético de RILs**. Viçosa, 2005. 120p. Dissertação (Mestrado), Universidade Federal de Viçosa.
- SCHUSTER I., CRUZ C.D. **Estatística genômica aplicada a populações derivadas de cruzamentos**. Viçosa , MG. Editora UFV, 2004. 585p.
- STEEL R.G.D., TORRIE J.H. **Principles and Procedures of Statistics: a Biometrical Approach**, McGraw-Hill, New York pp. 401-437., 1980.
- VARSHNEY R.K., GRANER A., SORRELLS M.E., Genic microsatellite markers in plants: features and applications. **Trends in Biotechnology**, v.23, n.1, p.48-55, 2005.
- VIJAYALAKSHMI S., YADAV K., KUSHWAHA C., SARODE S.B., SRIVASTAVA C.P., CHAND R., SINGH B.D. Identification of RAPD markers linked to the rust (*Uromyces fabae*) vresistance gene in pea (*Pisum sativum*). **Euphytica**, v.144, p.265-274, 2005.
- VISSCHER, P. M., C. S. HALEY, AND S. A. KNOTT, Mapping QTLs for binary traits in backcross and F2 populations. **Genetical Research**, v.68, p.55–63, 1996.
- ZENG, Z.B. Precision mapping of quantitative trait loci. **Genetics**, v.136, p.1457-14, 1994.
- ZOBEL, B., JETT J. B. **Genetics of wood production**. Berlin: Springer-Verlag, 1995. 337p.

- WEIR B.S., **Genetic data analysis II**. Sinauer Associates, Sunderland, 1996, 445p.
- WILCOX P.L., AMERSON H.V., KUHLMAN E.G., LIU B., O'MALLEY D.M., SEDEROFF R.R. Detection of a major gene for resistance to fusiform rust disease in loblolly pine by genomic mapping. **Proceedings of the National Academy of Sciences**, v.93, p.3859-3864, 1996.
- YI, N., S. XU, V. GEORGE, AND D. B. ALLISON Mapping multiple quantitative trait loci for ordinal trait. **Behavior Genetics**, v.34, p. 3–15, 2004.
- YOUNG N. D. A cautiously optimistic vision for marker-assisted breeding. **Molecular breeding**. v.5, p.505-510, 1999.
- YOUSEF G. G., JUVIK J. A., Comparison of phenotypic and marker assisted selection for quantitative traits in sweet corn. **Crop Science**. v.41 p.645-655, 2001.
- YU K., PARK S. J., POYSA V. Marker assisted selection of common beans for resistance to common bacterial blight: efficacy and economics. **Plant breeding**, v.199, p.411-415, 2000.
- XU, S., W. R. ATCHLEY, Mapping quantitative trait loci for complex binary diseases using line crosses. **Genetics**, v.143, p.1417–1424, 1996.

Capítulo II

**Mapeamento de locos controladores de características oligogênicas de
manifestação binária, utilizando funções de máxima verossimilhança -
Populações de F2.**

1. RESUMO

A busca por métodos que permitam explorar ao máximo a informação genética produzida durante a recombinação continua sendo um dos principais objetivos dos estudos de mapeamento e detecção de QTLs. As populações utilizadas no melhoramento vegetal são as principais matérias-primas para estudos de genética em plantas, entre as quais, destacam-se as populações F2. Características oligogênicas se caracterizam por sua herança governada por genes de maior efeito e por sua importância para as espécies vegetais cultivadas, com destaque para a resposta de resistência das plantas às doenças. Métodos tradicionais que pressupõem normalidade e herança governada por múltiplos fatores não deveriam ser utilizados para mapeamento destas características de distribuição discreta e segregação governada pela interação entre locos e alelos. O objetivo deste trabalho é propor um método para mapeamento e detecção de locos controladores da expressão de características oligogênicas, OTLs (“Oligogenic Trait Loci”) em populações F2. Este método definido como M.M.C.O (Método de Mapeamento de Características Oligogênicas) utiliza funções de verossimilhança para obtenção de estimativas de recombinação “ r ” de melhor ajuste aos dados, a partir das probabilidades condicionais de ocorrência entre o loco marcador e a herança da característica oligogênica. Os resultados mostram que o método foi adequado para detecção de OTLs em populações de F2. Os OTLs foram corretamente caracterizados e posicionados no mapa em populações de tamanho limitado, compostas por até 200 indivíduos. O M.M.C.O. não necessita de informação prévia de ordenamento dos marcadores e não é influenciado pela ocorrência de OTLs em um mesmo grupo de ligação. O atendimento das pressuposições para análise das características oligogênicas, a maior acurácia das estimativas, obtidos a partir da informação contida em todas as classes genotípicas e a possibilidade de posicionar pontualmente o OTL fazem do M.M.C.O. uma estratégia de potencial para contribuir na interpretação da herança de características oligogênicas.

2. ABSTRACT

The investigation of methods which allow a better interpretation of the genetic information produced during the recombination is still one of the major objectives of QTL detection. Oligogenic traits are distinguished by their heritage ruled by higher effect genes and by their importance for the cultivated plants, with emphasis to the plant disease resistance heritage. The qualitative nature and epistatic interaction of these traits results in a heritage pattern that should not be interpreted through traditional methods for QTL (Quantitative Trait Loci) detection. The objective of this work is to propose a method for detection of important loci related with the expression of oligogenic traits – OTL (Oligogenic Trait Loci). This method, defined as Oligogenic Trait Mapping Method (O.T.M.M.), uses maximum likelihood probability functions to obtain the r estimates of better adjustment to the observed data from the conditional probabilities of occurrence between the molecular markers and the OTL. The results show that the method was adequate for OTL detection in F₂ populations. The necessity of previous knowledge about the heritage pattern of the traits is one of the main limitations in the study of oligogenic traits, and the not proper definition of the correct heritage pattern results in a decrease in the accuracy of the methodology. The O.T.M.M. does not need previous information of the markers order in the linkage groups and are not influenced by the presence of two QTL in the same linkage group. The focus in the analysis presumes of the oligogenic heritage, the higher accuracy of the r estimates, obtained from the information contained in all genotypic classes, and the position of the OTL in the map, contribute to make O.T.M.M. a powerful tool for the analysis of oligogenic traits.

3. INTRODUÇÃO

A qualidade experimental é o principal critério a ser considerado no planejamento de experimentos de mapeamento e detecção de QTLs na área vegetal (VARSHNEY et al., 2005, SCHUSTER & CRUZ, 2004, BADU et al. 2004, ROCHA et al., 2002, YOUNG, 1999). A utilização de novos métodos e estratégias para análise de mapeamento justifica-se quando a violação nas pressuposições das metodologias tradicionais compromete a confiabilidade dos resultados (LI, 2004, RAO & LI, 2000).

Características de expressão governadas por poucos fatores de maior efeito têm-se mostrado importantes na condução dos programas de melhoramento de várias culturas, com destaque para a resposta de resistência de plantas a algumas doenças (ALZATE-MARIN et al., 2005, VIJAYALAKSHMI et al., 2005; ALFENAS et al., 2004, JUNGHANS et al., 2003; YU et al., 2000). As características oligogênicas possuem padrão de herança mendeliana e podem ser alocadas em classes de expressão governadas por dois, três ou mais genes de maior efeito, podendo apresentar interações epistáticas entre genes ou alelos de um mesmo gene, resultando em um padrão de herança mais complexo do que o padrão de herança governado por apenas um gene (LI, 2004, CRUZ et al., 2004, FALCONER, 1987).

A natureza discreta da resposta destas características associada à interação epistática de múltiplos fatores resulta em um padrão de segregação que não deveria ser mapeada pelos métodos tradicionais, que pressupõem normalidade e herança governada por muitos genes de pequeno efeito (LI, 2004, ZENG, 1994, JANSEN, 1993; LANDER & BOTSTEIN, 1989). As particularidades do mapeamento de características oligogênicas é assunto pertinente, abordado na literatura em vários trabalhos (YI et al., 2004, RAO e LI, 2000, VISSCHER et al., 1996, XU & ATCHLEY, 1996). VISSCHER et al.(1996) compararam métodos, utilizando regressão linear e regressão linear generalizada a partir de dados binários, e concluíram não haver diferença na acurácia da predição e no poder de teste entre as metodologias. XU & ATCHLEY (1996) estenderam a metodologia de

mapeamento por intervalo composto para dados binários utilizando modelos de regressão logística. RAO & LI (2000) apresentaram estratégias para análise de variáveis discretas utilizando diferentes modelos alélicos. E ainda existem autores que apresentam aproximações Bayesianas para detecção do QTL (YI et al., 2004). Contudo, nenhuma das estratégias pesquisadas aborda a questão da maneira considerada neste trabalho.

Uma das considerações mais importantes no estudo das características de distribuição discreta é o teste de hipótese genética que permite com uma margem de erro, concluir a respeito do padrão predominante de segregação, se governado por um ou mais genes e o tipo de interação predominante (SCHUSTER & CRUZ, 2004, LIU, 1997). Diferentemente da abordagem métrica, que suporta desvios em relação ao não atendimento das pressuposições, tais como aditividade de efeitos, distribuição normal, ausência de efeito epistático significativo e homogeneidade de erros (CRUZ et al., 2003, FALCONER, 1987), a consideração de uma hipótese genética equivocada resulta em interpretações com potencial para prejudicar a continuidade da pesquisa e procedimentos de seleção. Conclusões incorretas em relação à natureza da herança, ao genótipo das matrizes, ao posicionamento equivocado de locos nos mapas de ligação são alguns tipos de erro com potencial para gerar prejuízos.

Para a realização dos estudos de herança, deve-se ainda considerar que as populações utilizadas nos programas de melhoramento vegetal são, em muitos casos, a matéria-prima que o pesquisador deve explorar para seus estudos de genética (MURO ABAD et al., 2005a, MURO ABAD et al. 2005b, ROCHA et al., 2004). À disposição do pesquisador existem várias configurações de cruzamentos controlados que podem ser utilizados para aceitar com maior confiabilidade uma hipótese genética, sendo que entre os diferentes tipos de cruzamento e graus de parentesco, têm se destacado as populações F₂ (BOREM, 2001; CRUZ et al., 2003; CRUZ et al., 2004, BUNYAMIN et al., 2002, LYNCH & WALSH, 1998).

As populações F₂ são obtidas pela autofecundação das plantas F₁ resultantes do cruzamento entre genitores homozigóticos ou heterozigóticos. A necessidade de poucas gerações para desenvolvimento, a disponibilidade de 3

classes genotípicas para avaliação e a possibilidade de obtenção de natureza aditiva e de dominância separadamente justificam a ampla utilização destas populações para o mapeamento (SILVA, 2005; LIU, 1997). A utilização de plantas individuais para realização de medições talvez seja uma das suas principais limitações, resultando em uma diminuição na precisão das estimativas quando se compara com a utilização de réplicas (HARUSHIMA et al. 1998).

Assim, o objetivo deste trabalho é apresentar uma nova abordagem para mapeamento e detecção de locos controladores da expressão de características controladas por poucos genes de maior efeito com ênfase em segregações e interações epistáticas governadas por dois e três genes em populações F2. Visando ilustrar a aplicação e utilidade da metodologia proposta, foram simuladas populações F2 compostas por 200 irmãos completos, saturação média de cinco cM e seis características oligogênicas de expressão binária, que representam os padrões de segregação mais comuns observados nos programas de melhoramento e no mapeamento de espécies vegetais.

A utilização de populações simuladas permite quantificar a capacidade da metodologia proposta em recuperar a informação paramétrica por meio do conhecimento da condição referência, denominada por alguns autores como condição ou genoma paramétrico (SCHUSTER & CRUZ, 2004, SILVA, 2005) para avaliar o mérito e o poder de teste da metodologia para a detecção de locos mais importantes na expressão de características oligogênicas e na expressão binária (RAO & LI, 2000).

4. MATERIAL E MÉTODOS

4.1 Simulação dos dados

Foi utilizado o módulo de simulação do aplicativo computacional GQMOL (CRUZ, 2007) para a obtenção de populações F2 de uma espécie diplóide com 4 grupos de ligação de 200 cM de tamanho cada um, saturação média de 5 cM e marcadores de natureza co-dominante.

4.1.1 Simulação do genoma

Visando à avaliação de indivíduos recombinantes provenientes de uma configuração genômica mais parcimoniosa para as análises de recuperação da informação biológica paramétrica, foi considerado um genoma constituído de quatro grupos de ligação com 20 marcas eqüidistantes, co-dominantes e nível de saturação médio de 5 cM (Figura 1).

4.1.2 Simulação de genitores

Todas as populações simuladas foram obtidas por eventos de recombinação provenientes do mesmo arranjo de genes na população F1, sendo que o contraste entre o polimorfismo dos genitores foi resultado do cruzamento entre genitores contrastantes, situação esta que produz eventos de recombinação somente em fase de acoplamento.

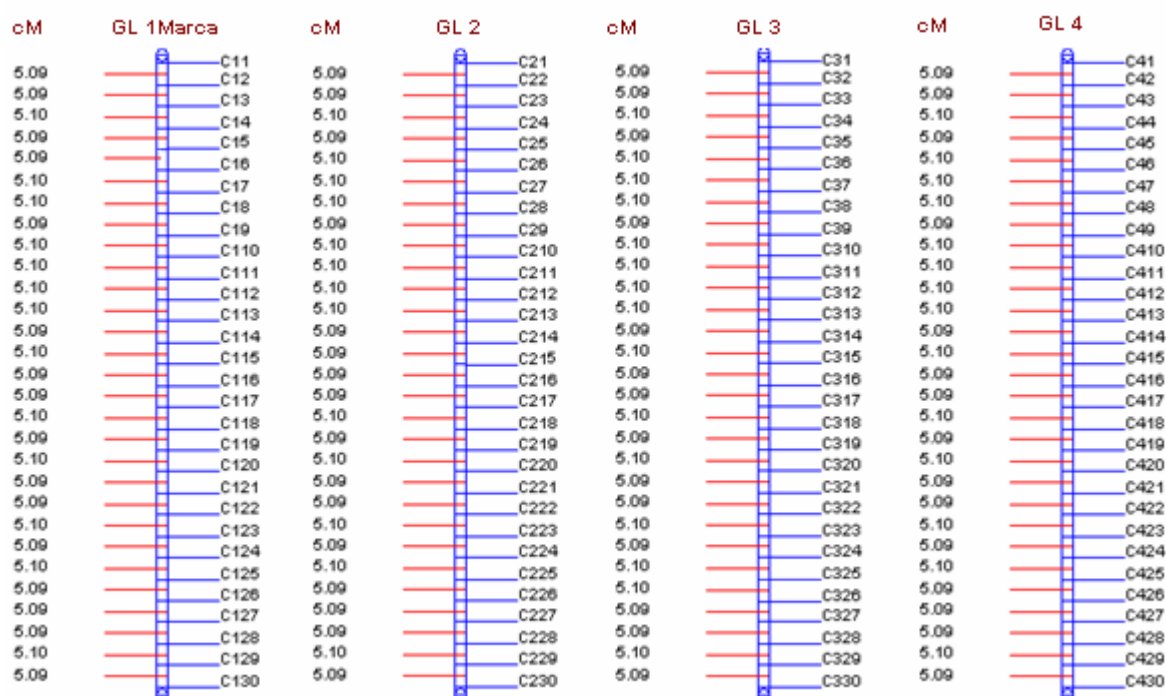


Figura 1 – Genoma com 4 grupos de ligação (GL) e saturação média de 5 cM, contendo 29 marcas nos grupos de ligação 1, 2 e 3 e 30 marcas no grupo de ligação 4, perfazendo um total de 117 marcas.

4.1.3 Simulação da característica oligogênica

Seis características governadas por dois e três genes de interação epistática predominantemente dominantes com proporções esperadas entre resistentes (1) e suscetíveis (0) de 9:7, 13:3, 15:1, 63:1 e 27:37 foram simuladas como mostrado na Tabela 1. Para simulação da epistasia entre os genes denominados de A, B e C foi considerado o seguinte posicionamento no genoma original: A (grupo de ligação 1, marcador 5), B (grupo de ligação 2, marcador 15) e C (grupo de ligação 3, marcador 26), sendo que estas foram responsáveis pela expressão de uma única variável que representa a segregação simultânea de dois ou três marcadores de acordo com as considerações em relação aos mecanismos de herança, denominadas condições, apresentados nas Tabelas 1 e 2.

Tabela 1- Descrição dos padrões de herança considerados para estudo

Condições	Descrição do padrão de herança
1	Característica governada por 2 genes, técnica de marcador molecular de natureza dominante, configuração de aproximação entre loco marcador e loco A.
2	Característica governada por 2 genes, técnica de marcador molecular de natureza dominante, configuração de repulsão entre loco marcador e loco A.
3	Característica governada por 2 genes, técnica de marcador molecular de natureza co-dominante, configuração de aproximação ou repulsão entre loco marcador e loco A.
4	Característica governada por 3 genes, técnica de marcador molecular de natureza dominante, configuração de aproximação entre loco marcador e loco A.
5	Característica governada por 3 genes, técnica de marcador molecular de natureza dominante, configuração de repulsão entre loco marcador e loco A.
6	Característica governada por 3 genes, técnica de marcador molecular de natureza co-dominante, configuração de aproximação ou repulsão entre loco marcador e loco A.

Tabela 2 – Classes genótipos e proporções esperadas considerando a epistasia do tipo dominante dos alelos A, B e C dos genes controladores das condições de herança avaliadas.

Interações epistáticas 2 genes							
9	A_B_	13	A_B_	13	A_B_	15	A_B_
7	A_bb		aabb		aabb		A_bb
	aaB_		A_bb		aaB_		aaB_
	aabb	3	aaB_	3	A_bb	1	aabb
Interações epistáticas 3 genes							
63	A_B_C_	27	A_B_C_				
	A_B_cc		A_B_cc				
	A_bbC_		A_bbC_				
	aaB_C_		aaB_C_				
	AAbbcc	37	AAbbcc				
	aaBBcc		aaBBcc				
	aabbCC		aabbCC				
1	aabbcc		aabbcc				

4.2 Análises genômicas

4.2.1. Mapeamento genético

No intuito de avaliar o método proposto para recuperação da informação de herança de características oligogênicas e manifestação binária, as análises descritas a seguir foram consideradas:

4.2.2 Análise de segregação de locos individuais – Teste de (χ^2)

Foi utilizado o teste de qui-quadrado (χ^2) para verificar a razão de segregação de cada marca em todas as populações geradas. A estatística qui-quadrado é dada por:

$$(i) \quad \chi^2_{\text{calculado}} = \sum_i \left[\frac{(O_i - E_i)^2}{E_i} \right]$$

4.2.3 - Análise de pares de marcas - estimação da porcentagem de recombinação

As estimativas de recombinação foram obtidas pelo método analítico definido por SCHUSTER & CRUZ (2004) que consistem da utilização da função de verossimilhança para a estimação dos valores de distância do mapa.

$$(ii) \quad L(p_i / n_i) = \lambda p_1^{n_1} p_2^{n_2} p_3^{n_3} \dots p_n^{n_n}$$

Em que :

$$(iii) \quad \lambda = \frac{N!}{n_1! n_2! \dots n_n!}$$

$$(iv) \quad N = \sum_{i=1}^n n_i$$

Em que:

N é o tamanho da população

n_i é o número de ocorrências do evento i com probabilidade p_i .

4.2.4 - Análise de pares de marcas - determinação dos grupos de ligação e ordenamento das marcas

A determinação dos grupos de ligação foi obtida pelo agrupamento baseado na frequência máxima de recombinação ($r_{\max}$) e o LOD mínimo ($LOD_{\min}$). Neste trabalho foram considerados os valores de $r_{\max} = 30\%$ e $LOD_{\min} = 3$. Após a formação de um grupo de ligação preliminar, verificou-se a melhor ordem dos n marcadores constituintes do grupo pelo método SARF (*“Sum of Adjacent Recombination Fractions”*).

4.2.5 Teste de segregação ambígua

LIU (1997) apresenta uma metodologia desenvolvida por Batley (1961) que permite a estimação do número de eventos necessários para a discriminação de

duas hipóteses genéticas considerando a distribuição de χ^2 e um nível crítico associado. Este estimador também apresentado por SCHUSTER & CRUZ(2004) considera dois padrões de segregação diferentes, definidos por x e y , e a frequência observada para as classes definidas por a e b . A falta de ajuste mensurada pelo teste do χ^2 é definida da seguinte maneira:

$$n \geq \left[\frac{1 + \sqrt{(xy)}}{\sqrt{x} + \sqrt{y}} \right]^2 \chi_{\alpha,1}^2$$

x frequência observada da classe a

y frequência observada da classe b

4.3 Comparação de genomas

4.3.1 Detecção de características oligogênicas.

Funções de verossimilhança foram utilizadas para obtenção de estimativas de distância entre os marcadores e os locos controladores da expressão destas características. O teste estatístico para testar a ligação entre o marcador e a característica oligogênica é o da razão de verossimilhança que consiste na comparação entre o estimador de máxima verossimilhança θ_A e a probabilidade de ocorrência de θ sob a hipótese nula de que $\theta_N = 0.5$; sendo 0.5 a distância entre o marcador e loco controlador da característica oligogênica.

(v)
$$L \frac{(\theta_A / x)}{(\theta_N / x)}$$

Para facilidade de interpretação, muitos autores preferem utilizar o logaritmo na base 10 da razão de verossimilhança.

$$(vi) \quad Z = \text{Log}_{10} \left[L \frac{(\theta_A / x)}{(\theta_N / x)} \right]$$

O valor da estimativa de Z significa que a hipótese alternativa é 10^Z vezes mais provável que a hipótese nula. A consistência, normalidade e eficiência assintótica são propriedades desejadas e que caracterizam os estimadores de máxima verossimilhança. A variância dos estimadores pode ser estimada pelo inverso do índice de informatividade.

5. RESULTADOS E DISCUSSÃO

5.1- Simulação dos dados

A escolha da ploidia referência deveu-se ao fato de as espécies diplóides serem representativas das espécies vegetais cultivadas (BOREM, 2001). A natureza co-dominante do marcador foi escolhida considerando a natureza das técnicas mais recentes utilizadas para mapeamento (BUNYAMIN et al., 2002, DEKKERS & HOSPITAL, 2001; BRONDANI et al., 1998, CHARLESWORTH et al., 1994). Quatro grupos de ligação são suficientes para explorar os fenômenos de epistasia que serão estudados e o valor de saturação escolhido é um valor conservador observado nos principais trabalhos de mapeamento. Um mapa com este valor de saturação é avaliado como medianamente saturado (SILVA, 2005). Muitos trabalhos sugerem que o número de 200 indivíduos deve ser considerado mínimo para o desenvolvimento de mapas de ligação (SILVA, 2005, SCHUSTER & CRUZ, 2004, MACKAY, 2001).

Foram simuladas populações F2 compostas por 200 irmãos completos, saturação média de 5 cM e duas características oligogênicas de expressão binária e natureza epistática, C1 e C2, governadas, respectivamente, pelas marcas 5 (G.L.1) e 14 (G.L.2) e pelas marcas 5 (G.L.1), 14 (G.L.2) e 26 (G.L.3). As condições deste estudo de simulação foram definidas visando obter conclusões que sejam aplicadas, tendo em vista que muitas pesquisas de mapeamento

realizadas com 200 indivíduos e saturação média de 5 cM almejam o desenvolvimento de mapas referências (NOVAES, 2006, MISSIAGIA, 2005, MEKSEM et al., 2001, CERVERA et al., 2001, SILVA, 2005,).

5.2 – Estimativas de probabilidade condicional

A utilização da função de verossimilhança permite a obtenção de estimativas de variância mínima que são estimadas a partir da informação contida em todas as classes genótípicas que compõem a população, (SCHUSTER & CRUZ, 2004, LYNCH & WALSH, 1998). As estimativas de probabilidade condicional expressam a razão da ocorrência conjunta das classes genótípicas do QTL e do marcador e são necessárias para a estimação dos valores de recombinação utilizando o método de máxima verossimilhança (SCHUSTER & CRUZ, 2004, LIU, 1997). As estimativas de probabilidade apresentadas a seguir consideram as pressuposições em relação ao padrão de herança: i) expressão binária do tipo presença ou ausência e ii) herança governada por poucos genes e segregação correlacionada com uma marca molecular previamente mapeada.

A inclusão de outros valores de probabilidade condicional associados a padrões de segregação específicos permite adaptar as funções de verossimilhança a quaisquer outros tipos de populações e características. O maior número de cálculos necessários para a obtenção da resolução destas funções é a principal limitação para o uso desta ferramenta (WEIR, 1996; LIU, 1997)

A seguir são apresentadas as estimativas de probabilidade de ocorrência das classes genótípicas em relação ao loco marcador (M) e aos locos controladores das características (A e B, ou A, B e C) considerando: herança governada por 2 ou por 3 genes, tipos de interação epistáticas mais comuns, fase de ligação e a natureza da técnica de marcador molecular utilizada se dominante ou co-dominante, resultando em um total de seis condições. Tais estimadores de probabilidade da ocorrência individual dos locos permitem a obtenção das estimativas de probabilidades de segregação condicional entre o marcador e os locos controladores da resistência.

Condição 1: Característica governada por 2 genes, técnica de marcador molecular de natureza dominante, configuração de aproximação entre loco marcador (M) e loco A.

Tabela 4- Proporção teórica esperada das classes genotípicas em população F2 com padrões de segregação esperado de 9:7, 13:3a, 13:3b, 15:1 estimadas em função de r .

Classes genotípicas		Probabilidade em função de r	Classes Fenotípicas 9:7	Classes Fenotípicas 13:3a	Classes Fenotípicas 13:3b	Classes Fenotípicas 15:1
M_A_	B_	$\frac{3}{4} \left[\frac{1}{2} + \frac{1}{4}(1-r)^2 \right]$	1R	1R	1R	1R
	bb	$\frac{1}{4} \left[\frac{1}{2} + \frac{1}{4}(1-r)^2 \right]$	1S	1S	1R	1R
M_aa	B_	$\frac{3}{4} \left[\frac{1}{4} - \frac{1}{4}(1-r)^2 \right]$	1S	1R	1S	1R
	bb	$\frac{1}{4} \left[\frac{1}{4} - \frac{1}{4}(1-r)^2 \right]$	1S	1R	1R	1S
mmA_	B_	$\frac{3}{4} \left[\frac{1}{4} - \frac{1}{4}(1-r)^2 \right]$	0R	0R	0R	0R
	bb	$\frac{1}{4} \left[\frac{1}{4} - \frac{1}{4}(1-r)^2 \right]$	0S	0S	0R	0R
mmaa	B_	$\frac{3}{4} \left[\frac{1}{4}(1-r)^2 \right]$	0S	0R	0S	0R
	bb	$\frac{1}{4} \left[\frac{1}{4}(1-r)^2 \right]$	0S	0R	0R	0S

Tabela 5- Frequências esperadas das classes genotípicas em função de r

Classes genotípicas	Frequências esperadas em função de r 9:7	Frequências esperadas em função de r 13:3a	Frequências esperadas em função de r 13:3b	Frequências esperadas em função de r 15:1
1R	$\frac{3}{8} + \frac{3}{16}s^2$	$\frac{10}{16} - \frac{1}{16}s^2$	$\frac{9}{16} + \frac{3}{16}s^2$	$\frac{11}{16} + \frac{1}{16}s^2$
1S	$\frac{6}{16} - \frac{3}{16}s^2$	$\frac{2}{16} + \frac{1}{16}s^2$	$\frac{3}{16} - \frac{3}{16}s^2$	$\frac{1}{16} - \frac{1}{16}s^2$
0R	$\frac{3}{16} - \frac{3}{16}s^2$	$\frac{3}{16} + \frac{1}{16}s^2$	$\frac{4}{16} - \frac{3}{16}s^2$	$\frac{4}{16} - \frac{1}{16}s^2$
0S	$\frac{1}{16} + \frac{3}{16}s^2$	$\frac{1}{16} - \frac{1}{16}s^2$	$\frac{3}{16}s^2$	$\frac{1}{16}s^2$

Em que $s = (1-r)$

Condição 2: Característica governada por 2 genes, técnica de marcador molecular de natureza dominante, configuração de repulsão entre loco marcador (M) e loco A.

Tabela 6- Proporção teórica esperada das classes genotípicas em população F2 com padrões de segregação esperado de 9:7, 13:3a, 13:3b, 15:1 estimadas em função de r .

Classes genotípicas		Freqüências esperadas em função de r	Classes Fenotípicas 9:7	Classes Fenotípicas 13:3a	Classes Fenotípicas 13:3b	Classes Fenotípicas 15:1
M_A_	B_	$3/4 [1/2 + 1/4r^2]$	1R	1R	1R	1R
	bb	$1/4 [1/2 + 1/4r^2]$	1S	1S	1R	1R
M_aa	B_	$3/4 [1/4 - 1/4r^2]$	1S	1R	1S	1R
	bb	$1/4 [1/4 - 1/4r^2]$	1S	1R	1R	1S
mmA_	B_	$3/4 [1/4 - 1/4r^2]$	0R	0R	0R	0R
	bb	$1/4 [1/4 - 1/4r^2]$	0S	0S	0R	0R
mmaa	B_	$3/4r^2$	0S	0R	0S	0R
	bb	$1/4r^2$	0S	0R	0R	0S

Tabela 7- Freqüências esperadas das classes genotípicas em função de r .

Classes genotípicas	Freqüências esperadas em função de r 9:7	Freqüências esperadas em função de r 13:3a	Freqüências esperadas em função de r 13:3b	Freqüências esperadas em função de r 15:1
1R	$\frac{6}{16} + \frac{3}{4}s$	$\frac{10}{16} - \frac{1}{4}s$	$\frac{9}{16} + \frac{3}{4}s$	$\frac{11}{16} + \frac{1}{4}s$
1S	$\frac{6}{16} - \frac{3}{4}s$	$\frac{2}{16} + \frac{1}{4}s$	$\frac{3}{16} - \frac{3}{4}s$	$\frac{1}{16} - \frac{1}{4}s$
0R	$\frac{3}{16} - \frac{3}{4}s$	$\frac{3}{16} + \frac{1}{4}s$	$\frac{4}{16} - \frac{3}{4}s$	$\frac{4}{16} - \frac{1}{4}s$
0S	$\frac{1}{16} + \frac{3}{4}s$	$\frac{1}{16} - \frac{1}{4}s$	$\frac{3}{4}s$	$\frac{1}{4}s$

Em que $s = \frac{1}{4}r^2$

Condição 3: Característica governada por 2 genes, técnica de marcador molecular de natureza co-dominante, configuração de aproximação entre loco marcador (M) e loco A.

Tabela 8- Proporção teórica esperada das classes genotípicas em população F2 com padrões de segregação esperado de 9:7, 13:3a, 13:3b, 15:1.

Classes genotípicas		Frequências esperadas em função de r	Classes Fenotípicas 9:7	Classes Fenotípicas 13:3a	Classes Fenotípicas 13:3b	Classes Fenotípicas 15:1
MMA_	B_	$\frac{3}{4} \left[\frac{1}{4} (1 - r^2) \right]$	2R	2R	2R	2R
	bb	$\frac{1}{4} \left[\frac{1}{4} (1 - r^2) \right]$	2S	2S	2S	2S
MmA_	B_	$\frac{3}{4} \left[\frac{1}{2} (1 - r + r^2) \right]$	1R	1R	1R	1R
	bb	$\frac{1}{4} \left[\frac{1}{2} (1 - r + r^2) \right]$	1S	1S	1R	1R
mmA_	B_	$\frac{3}{4} \left[\frac{1}{4} r (2 - r) \right]$	0R	0R	0R	0R
	bb	$\frac{1}{4} \left[\frac{1}{4} r (2 - r) \right]$	0S	0S	0R	0R
MMaa	B_	$\frac{3}{4} \left[\frac{1}{4} r^2 \right]$	2S	2R	2S	2R
	bb	$\frac{1}{4} \left[\frac{1}{4} r^2 \right]$	2S	2R	2R	2S
Mmaa	B_	$\frac{3}{4} \left[\frac{1}{2} r (1 - r) \right]$	1S	1R	1S	1R
	bb	$\frac{1}{4} \left[\frac{1}{2} r (1 - r) \right]$	1S	1R	1R	1S
mmaa	B_	$\frac{3}{4} \left[\frac{1}{4} (1 - r^2) \right]$	0S	1S	0S	0R
	bb	$\frac{1}{4} \left[\frac{1}{4} (1 - r^2) \right]$	0S	1S	0R	0S

Tabela 9- Frequências esperadas das classes genotípicas em função de r.

Classes genotípicas		Frequências esperadas em função de r 9:7	Frequências esperadas em função de r 13:3a	Frequências esperadas em função de r 13:3b	Frequências esperadas em função de r 15:1
2	R	$\frac{3}{16} - \frac{3}{16} r^2$	$\frac{3}{16} + \frac{1}{16} r^2$	$\frac{4}{16} - \frac{3}{16} r^2$	$\frac{4}{16} - \frac{1}{16} r^2$
	S	$\frac{1}{16} + \frac{3}{16} r^2$	$\frac{1}{16} - \frac{1}{16} r^2$	$\frac{3}{16} r^2$	$\frac{1}{16} r^2$
1	R	$\frac{6}{16} + \frac{6}{16} r^2 - \frac{6}{16} r$	$\frac{6}{16} - \frac{2}{16} r^2 + \frac{2}{16} r$	$\frac{8}{16} + \frac{6}{16} r^2 - \frac{6}{16} r$	$\frac{8}{16} + \frac{2}{16} r^2 - \frac{2}{16} r$
	S	$\frac{2}{16} - \frac{6}{16} r^2 + \frac{6}{16} r$	$\frac{2}{16} + \frac{2}{16} r^2 - \frac{2}{16} r$	$-\frac{6}{16} r^2 + \frac{6}{16} r$	$-\frac{2}{16} r^2 + \frac{2}{16} r$
0	R	$-\frac{3}{16} r^2 + \frac{6}{16} r^2$	$\frac{4}{16} + \frac{1}{16} r^2 - \frac{2}{16} r$	$\frac{1}{16} - \frac{3}{16} r^2 + \frac{6}{16} r$	$\frac{3}{16} - \frac{1}{16} r^2 + \frac{2}{16} r$
	S	$\frac{4}{16} + \frac{3}{16} r^2 - \frac{6}{16} r$	$-\frac{1}{16} r^2 + \frac{2}{16} r$	$\frac{3}{16} + \frac{3}{16} r^2 - \frac{6}{16} r$	$\frac{1}{16} + \frac{1}{16} r^2 - \frac{2}{16} r$

Condição 4: Característica governada por 3 genes, técnica de marcador molecular de natureza dominante, configuração de aproximação entre loco marcador (M) e loco A.

Tabela 10- Proporção teórica esperada das classes genotípicas em população F2 com padrões de segregação esperados de 9:7, 13:3a, 13:3b, 15:1 estimados em função de r.

Classes genotípicas			Freqüências esperadas em função de r	Classes Fenotípicas 63:1	Classes Fenotípicas 27:37
M_A_	B_ C_	9/16	$1/2 + 1/4(1-r)^2$	1R	1R
	B_ Cc	3/16	$1/2 + 1/4(1-r)^2$	1R	1S
	bb C_	3/16	$1/2 + 1/4(1-r)^2$	1R	1S
	bb Cc	1/16	$1/2 + 1/4(1-r)^2$	1R	1S
M_aa	B_ C_	9/16	$1/4 - 1/4(1-r)^2$	1R	1S
	B_ Cc	3/16	$1/4 - 1/4(1-r)^2$	1R	1S
	bb C_	3/16	$1/4 - 1/4(1-r)^2$	1R	1S
	bb Cc	1/16	$1/4 - 1/4(1-r)^2$	1S	1S
mmA_	B_ C_	9/16	$1/4 - 1/4(1-r)^2$	0R	0R
	B_ Cc	3/16	$1/4 - 1/4(1-r)^2$	0R	0S
	bb C_	3/16	$1/4 - 1/4(1-r)^2$	0R	0S
	bb Cc	1/16	$1/4 - 1/4(1-r)^2$	0R	0S
mmaa	B_ C_	9/16	$1/4(1-r)^2$	0R	0S
	B_ Cc	3/16	$1/4(1-r)^2$	0R	0S
	bb C_	3/16	$1/4(1-r)^2$	0R	0S
	bb Cc	1/16	$1/4(1-r)^2$	0S	0S

Condição 4: Característica governada por 3 genes, técnica de marcador molecular de natureza dominante, configuração de aproximação entre loco marcador (M) e o loco A.

Tabela 11- Frequências esperadas das classes genotípicas em função de r .

Classes genotípicas	Frequências esperadas em função de r	Frequências esperadas em função de r
	63:1	27:37
1R	$\frac{47}{64} + \frac{1}{16}s$	$\frac{18}{64} + \frac{9}{16}s$
1S	$\frac{1}{64} - \frac{1}{16}s$	$\frac{30}{64} - \frac{9}{16}s$
0R	$\frac{16}{64} - \frac{1}{16}s$	$\frac{9}{64} - \frac{9}{16}s$
0S	$\frac{1}{16}s$	$\frac{7}{64} + \frac{9}{16}s$

Em que $s = 1/4(1-r)^2$

Condição 5: Característica governada por 3 genes, técnica de marcador molecular de natureza dominante, configuração de repulsão entre loco marcador e loco A.

Tabela 12- Proporção teórica esperada das classes genotípicas em população F2 com padrões de segregação esperados de 9:7, 13:3a, 13:3b, 15:1 estimados em função de r.

Classes genotípicas			Frequências esperadas em função de r	Classes Fenotípicas 63:1	Classes Fenotípicas 27:37
M_A_	B_	C_	$9/16[1/2 + 1/4r^2]$	1R	1R
	B_	cc	$3/16[1/2 + 1/4r^2]$	1R	1S
	bb	C_	$3/16[1/2 + 1/4r^2]$	1R	1S
	bb	cc	$1/16[1/2 + 1/4r^2]$	1R	1S
M_aa	B_	C_	$9/16[1/4 - 1/4r^2]$	1R	1S
	B_	cc	$3/16[1/4 - 1/4r^2]$	1R	1S
	bb	C_	$3/16[1/4 - 1/4r^2]$	1R	1S
	bb	Cc	$1/16[1/4 - 1/4r^2]$	1S	1S
mmA_	B_	C_	$9/16[1/4 - 1/4r^2]$	0R	0R
	B_	cc	$3/16[1/4 - 1/4r^2]$	0R	0S
	bb	C_	$3/16[1/4 - 1/4r^2]$	0R	0S
	bb	cc	$1/16[1/4 - 1/4r^2]$	0R	0S
mmaa	B_	C_	$9/16(r^2)$	0R	0S
	B_	cc	$3/16(r^2)$	0R	0S
	bb	C_	$3/16(r^2)$	0R	0S
	bb	cc	$1/16(r^2)$	0S	0S

Condição 5: Característica governada por 3 genes, técnica de marcador molecular de natureza dominante, configuração de repulsão entre loco marcador e loco A.

Tabela 13- Freqüências esperadas das classes genotípicas em função de r .

Classes genotípicas	Freqüências esperadas em função de r	Freqüências esperadas em função de r
	63:1	27:37
1R	$\frac{47}{64} + \frac{1}{16}s$	$\frac{18}{64} + \frac{9}{16}s$
1S	$\frac{1}{64} - \frac{1}{16}s$	$\frac{30}{64} - \frac{9}{16}s$
0R	$\frac{16}{64} - \frac{1}{16}s$	$\frac{9}{64} - \frac{9}{16}s$
0S	$\frac{1}{16}s$	$\frac{7}{64} + \frac{9}{16}s$

Em que $s = 1/4r^2$

Condição 6: Característica governada por 3 genes, técnica de marcador molecular de natureza co-dominante, configuração de aproximação entre loco marcador (M) e loco A.

Tabela 14- Proporção teórica esperada das classes genotípicas em população F2 com padrões de segregação esperados de 9:7, 13:3a, 13:3b, 15:1 estimados em função de r.

Classes genotípicas			Frequências esperadas em função de r	Classes Fenotípicas 63:1	Classes Fenotípicas 27:37
M_A_	B_	C_	$\frac{3}{4} \left[\frac{1}{4} (1 - r^2) \right]$	2R	2R
	B_	cc	$\frac{3}{4} \left[\frac{1}{4} (1 - r^2) \right]$	2R	2S
	bb	C_	$\frac{3}{4} \left[\frac{1}{4} (1 - r^2) \right]$	2R	2S
	bb	cc	$\frac{3}{4} \left[\frac{1}{4} (1 - r^2) \right]$	2R	2S
MmA_	B_	C_	$\frac{3}{4} \left[\frac{1}{2} (1 - r + r^2) \right]$	1R	1R
	B_	cc	$\frac{3}{4} \left[\frac{1}{2} (1 - r + r^2) \right]$	1R	1S
	bb	C_	$\frac{3}{4} \left[\frac{1}{2} (1 - r + r^2) \right]$	1R	1S
	bb	Cc	$\frac{3}{4} \left[\frac{1}{2} (1 - r + r^2) \right]$	1R	1S
mmA_	B_	C_	$\frac{3}{4} \left[\frac{1}{4} r (2 - r) \right]$	0R	0R
	B_	cc	$\frac{3}{4} \left[\frac{1}{4} r (2 - r) \right]$	0R	0S
	bb	C_	$\frac{3}{4} \left[\frac{1}{4} r (2 - r) \right]$	0R	0S
	bb	cc	$\frac{3}{4} \left[\frac{1}{4} r (2 - r) \right]$	0R	0S
MMaa	B_	C_	$\frac{3}{4} \left[\frac{1}{4} r^2 \right]$	2R	2S
	B_	cc	$\frac{3}{4} \left[\frac{1}{4} r^2 \right]$	2R	2S
	bb	C_	$\frac{3}{4} \left[\frac{1}{4} r^2 \right]$	2R	2S
	bb	cc	$\frac{3}{4} \left[\frac{1}{4} r^2 \right]$	2S	2S
Mmaa	B_	C_	$\frac{3}{4} \left[\frac{1}{2} r (1 - r) \right]$	1R	1S
	B_	cc	$\frac{3}{4} \left[\frac{1}{2} r (1 - r) \right]$	1R	1S
	bb	C_	$\frac{3}{4} \left[\frac{1}{2} r (1 - r) \right]$	1R	1S
	bb	cc	$\frac{3}{4} \left[\frac{1}{2} r (1 - r) \right]$	1S	1S
mmaa	B_	C_	$\frac{1}{4} \left[\frac{1}{4} (1 - r^2) \right]$	0R	0S
	B_	cc	$\frac{1}{4} \left[\frac{1}{4} (1 - r^2) \right]$	0R	0S
	bb	C_	$\frac{1}{4} \left[\frac{1}{4} (1 - r^2) \right]$	0R	0S
	bb	cc	$\frac{1}{4} \left[\frac{1}{4} (1 - r^2) \right]$	0S	0S

Condição 6: Característica governada por 3 genes, técnica de marcador molecular de natureza co-dominante, configuração de aproximação entre o loco marcador (M) e loco A.

Tabela 15- Freqüências esperadas das classes genotípicas em função de r .

Classes genotípicas		Freqüências esperadas em função de r	Freqüências esperadas em função de r
		63:1	27:37
2	R	$\frac{16}{64} - \frac{1}{64}r^2$	$\frac{9}{64} - \frac{9}{64}r^2$
	S	$\frac{1}{64}r^2$	$\frac{7}{64} + \frac{9}{64}r^2$
1	R	$\frac{32}{64} + \frac{2}{64}r^2 - \frac{2}{64}r$	$\frac{18}{64} + \frac{18}{64}r^2 - \frac{18}{64}r$
	S	$-\frac{2}{64}r^2 + \frac{2}{64}r$	$\frac{14}{64} - \frac{18}{64}r^2 + \frac{18}{64}r$
0	R	$\frac{15}{64} - \frac{1}{64}r^2 + \frac{2}{64}r$	$-\frac{9}{64}r^2 + \frac{18}{64}r$
	S	$\frac{1}{64} + \frac{1}{64}r^2 - \frac{2}{64}r$	$\frac{16}{64} + \frac{9}{64}r^2 - \frac{18}{64}r$

5.3- Funções de verossimilhança

Além das estimativas de probabilidade condicional, é necessário para estimação do valor de recombinação o conhecimento da função densidade de probabilidade (f.d.p.) da distribuição da variável aleatória (SCHUSTER & CRUZ, 2004, LIU, 1997; WEIR, 1996) . Desta maneira, considera-se que a probabilidade de ocorrência das classes genotípicas nas populações segregantes assumem distribuição multinomial (LIU, 1997). A primeira derivada da função suporte em relação ao parâmetro que se deseja estimar permite a obtenção da função escore (f'), que, por sua vez, permite estimar o ponto de “r” de máxima probabilidade de ajuste (SCHUSTER & CRUZ, 2004). As funções de verossimilhança utilizadas para estimar o valor de r entre marcadores e QTL para as características oligogênicas são apresentadas abaixo:

Condição 1:

Proporção esperada 9:7

$$L(p_i / n_i) = \lambda \left(\frac{3}{8} + \frac{3}{16} s^2 \right)^{n_1} \left(\frac{6}{16} - \frac{3}{16} s^2 \right)^{n_2} \left(\frac{3}{16} - \frac{3}{16} s^2 \right)^{n_3} \left(\frac{1}{16} + \frac{3}{16} s^2 \right)^{n_4}$$

Proporção esperada 13:3a

$$L(p_i / n_i) = \lambda \left(\frac{10}{16} - \frac{1}{16} s^2 \right)^{n_1} \left(\frac{2}{16} + \frac{1}{16} s^2 \right)^{n_2} \left(\frac{3}{16} + \frac{1}{16} s^2 \right)^{n_3} \left(\frac{1}{16} - \frac{1}{16} s^2 \right)^{n_4}$$

$$s = (1 - r)$$

Proporção esperada 13:3b

$$L(p_i / n_i) = \lambda \left(\frac{10}{16} - \frac{1}{16} s^2 \right)^{n1} \left(\frac{2}{16} + \frac{1}{16} s^2 \right)^{n2} \left(\frac{3}{16} + \frac{1}{16} s^2 \right)^{n3} \left(\frac{1}{16} - \frac{1}{16} s^2 \right)^{n4}$$

Proporção esperada 15:1

$$L(p_i / n_i) = \lambda \left(\frac{11}{16} + \frac{1}{16} s^2 \right)^{n1} \left(\frac{1}{16} - \frac{1}{16} s^2 \right)^{n2} \left(\frac{4}{16} - \frac{1}{16} s^2 \right)^{n3} \left(\frac{1}{16} s^2 \right)^{n4}$$

Condição 2:**Proporção esperada 9:7**

$$L(p_i / n_i) = \lambda \left(\frac{6}{16} + \frac{3}{4} s \right)^{n1} \left(\frac{6}{16} - \frac{3}{4} s \right)^{n2} \left(\frac{3}{16} - \frac{3}{4} s \right)^{n3} \left(\frac{1}{16} + \frac{3}{4} s \right)^{n4}$$

Proporção esperada 13:3a

$$L(p_i / n_i) = \lambda \left(\frac{10}{16} - \frac{1}{4} s \right)^{n1} \left(\frac{2}{16} + \frac{1}{4} s \right)^{n2} \left(\frac{3}{16} + \frac{1}{4} s \right)^{n3} \left(\frac{1}{16} - \frac{1}{4} s \right)^{n4}$$

Proporção esperada 13:3b

$$L(p_i / n_i) = \lambda \left(\frac{9}{16} + \frac{3}{4} s \right)^{n1} \left(\frac{3}{16} - \frac{3}{4} s \right)^{n2} \left(\frac{4}{16} - \frac{3}{4} s \right)^{n3} \left(\frac{3}{4} s \right)^{n4}$$

Proporção esperada 15:1

$$L(p_i / n_i) = \lambda \left(\frac{11}{16} + \frac{1}{4} s \right)^{n1} \left(\frac{1}{16} - \frac{1}{4} s \right)^{n2} \left(\frac{4}{16} - \frac{1}{4} s \right)^{n3} \left(\frac{1}{4} s \right)^{n4}$$

$$s = 1/4 r^2$$

Condição 3:**Proporção esperada 9:7**

$$L(p_i / n_i) = \lambda \left(\frac{3}{16} - \frac{3}{16} r^2 \right)^{n1} \left(\frac{1}{16} + \frac{3}{16} r^2 \right)^{n2} \left(\frac{6}{16} + \frac{6}{16} r^2 - \frac{6}{16} r \right)^{n3} \left(\frac{2}{16} - \frac{6}{16} r^2 + \frac{6}{16} r \right)^{n4}$$

$$\left(-\frac{3}{16} r^2 + \frac{6}{16} r^2 \right)^{n5} \left(\frac{4}{16} + \frac{3}{16} r^2 - \frac{6}{16} r \right)^{n6}$$

Proporção esperada 13:3a

$$L(p_i / n_i) = \lambda \left(\frac{3}{16} - \frac{3}{16} r^2 \right)^{n1} \left(\frac{1}{16} - \frac{1}{16} r^2 \right)^{n2} \left(\frac{6}{16} - \frac{2}{16} r^2 + \frac{2}{16} r \right)^{n3} \left(\frac{2}{16} + \frac{2}{16} r^2 - \frac{2}{16} r \right)^{n4} \\ \left(\frac{4}{16} + \frac{1}{16} r^2 - \frac{2}{16} r \right)^{n5} \left(-\frac{1}{16} r^2 + \frac{2}{16} r \right)^{n6}$$

Proporção esperada 13:3b

$$L(p_i / n_i) = \lambda \left(\frac{4}{16} - \frac{3}{16} r^2 \right)^{n1} \left(\frac{3}{16} r^2 \right)^{n2} \left(\frac{8}{16} + \frac{6}{16} r^2 - \frac{6}{16} r \right)^{n3} \left(-\frac{6}{16} r^2 + \frac{6}{16} r \right)^{n4} \\ \left(\frac{1}{16} - \frac{3}{16} r^2 + \frac{6}{16} r \right)^{n5} \left(\frac{3}{16} + \frac{3}{16} r^2 - \frac{6}{16} r \right)^{n6}$$

Proporção esperada 15:1

$$L(p_i / n_i) = \lambda \left(\frac{4}{16} - \frac{1}{16} r^2 \right)^{n1} \left(\frac{1}{16} r^2 \right)^{n2} \left(\frac{8}{16} + \frac{2}{16} r^2 - \frac{2}{16} r \right)^{n3} \left(-\frac{2}{16} r^2 + \frac{2}{16} r \right)^{n4} \\ \left(\frac{3}{16} - \frac{1}{16} r^2 + \frac{2}{16} r \right)^{n5} \left(\frac{1}{16} + \frac{1}{16} r^2 - \frac{2}{16} r \right)^{n6}$$

Condição 4:**Proporção esperada 63:1**

$$L(p_i / n_i) = \lambda \left(\frac{47}{64} + \frac{1}{16} s \right)^{n1} \left(\frac{1}{64} - \frac{1}{16} s \right)^{n2} \left(\frac{16}{64} - \frac{1}{16} s \right)^{n3} \left(\frac{1}{16} s \right)^{n4}$$

Proporção esperada 27:37

$$L(p_i / n_i) = \lambda \left(\frac{18}{64} + \frac{9}{16} s \right)^{n1} \left(\frac{30}{64} - \frac{9}{16} s \right)^{n2} \left(\frac{9}{64} - \frac{9}{16} s \right)^{n3} \left(\frac{7}{64} + \frac{9}{16} s \right)^{n4}$$

$$s = 1/4(1-r)^2$$

Condição 5:**Proporção esperada 63:1**

$$L(p_i / n_i) = \lambda \left(\frac{47}{64} + \frac{1}{16} s \right)^{n_1} \left(\frac{1}{64} - \frac{1}{16} s \right)^{n_2} \left(\frac{16}{64} - \frac{1}{16} s \right)^{n_3} \left(\frac{1}{16} s \right)^{n_4}$$

Proporção esperada 27:37

$$L(p_i / n_i) = \lambda \left(\frac{18}{64} + \frac{9}{16} s \right)^{n_1} \left(\frac{30}{64} - \frac{9}{16} s \right)^{n_2} \left(\frac{9}{64} - \frac{9}{16} s \right)^{n_3} \left(\frac{7}{64} + \frac{9}{16} s \right)^{n_4}$$

$$s = 1/4 r^2$$

Condição 6:**Proporção esperada 63:1**

$$L(p_i / n_i) = \lambda \left(\frac{16}{64} - \frac{1}{64} r^2 \right)^{n_1} \left(\frac{1}{64} r^2 \right)^{n_2} \left(\frac{32}{64} + \frac{2}{64} r^2 - \frac{2}{64} r \right)^{n_3} \left(-\frac{2}{64} r^2 + \frac{2}{64} r \right)^{n_4}$$

$$\left(\frac{15}{64} - \frac{1}{64} r^2 + \frac{2}{64} r \right)^{n_5} \left(\frac{1}{64} + \frac{1}{64} r^2 - \frac{2}{64} r \right)^{n_6}$$

Proporção esperada 27:37

$$L(p_i / n_i) = \lambda \left(\frac{9}{64} - \frac{9}{64} r^2 \right)^{n_1} \left(\frac{7}{64} + \frac{9}{64} r^2 \right)^{n_2} \left(\frac{18}{64} + \frac{18}{64} r^2 - \frac{18}{64} r \right)^{n_3} \left(\frac{14}{64} - \frac{18}{64} r^2 + \frac{18}{64} r \right)^{n_4}$$

$$\left(-\frac{9}{64} r^2 + \frac{18}{64} r \right)^{n_5} \left(\frac{16}{64} + \frac{9}{64} r^2 - \frac{18}{64} r \right)^{n_6}$$

Em que :

$$\lambda = \frac{N!}{n_1! n_2! \dots n_n!}$$

$$N = \sum_{i=1}^n n_i$$

5.4 – Mapeamento genético

Segundo LIU (1997), a obtenção de um mapa genético de qualidade é pré-requisito para a detecção de QTLs. Neste contexto, a interpretação dos dados de mapeamento foi realizada visando inferir a qualidade do mapa estimado a partir dos parâmetros definidos para simulação.

Em etapa anterior ao desenvolvimento dos mapas de ligação, as populações avaliadas tiveram sua segregação individual testada pelo teste de qui-quadrado com correção de Bonferroni para testes múltiplos. Do conjunto total de marcas avaliadas, apenas dois marcadores de populações distintas apresentaram segregação distorcida segundo o teste de qui-quadrado. Considerando o total de 720 marcadores analisados, espera-se, devido ao acaso ($\alpha < 0,01$), que 7,2 marcas apresentem distorção da proporção esperada proveniente da avaliação de n populações. No caso específico, em que a obtenção de eventos de recombinação é responsabilidade do acaso, foi feita a opção de não descartar os marcadores que apresentaram segregação distorcida.

Para a obtenção dos grupos de ligação, foram utilizados os parâmetros de “LOD score” mínimo (LOD = 3,0) e a máxima frequência de recombinação ($r = 0,30$), em comparação com os valores tradicionalmente considerados nos trabalhos de mapeamento. No entanto, observou-se que valores mais elevados de “LOD SCORE” (LOD = 30) poderiam ser considerados sem resultar em uma redução no número de marcas que compõem os grupos de ligação das populações avaliadas. Este valor de LOD é dez vezes mais elevado que o nível mínimo sugerido para rejeição da hipótese de segregação independente e indica que os parâmetros definidos para desenvolvimento das populações de F2 avaliadas permitem obtenção de estimativas de recombinação com precisão suficiente para posicionamento dos locos QTLs.

Informações “a priori” relativas à obtenção de estimativas de recombinação em populações de F2 utilizando de funções de máxima verossimilhança e baseadas em índices de informatividade indicam que esta população é uma das que conferem as estimativas mais precisas junto com as populações de

retrocruzamento e de duplo-haplóides (ROCHA et al., 2004, SCHUSTER & CRUZ, 2004). O desvio entre os valores paramétricos e os valores estimados em função dos dados observados também pode ser avaliado como medida de qualidade dos mapas genéticos considerados neste estudo (Tabela 16).

Tabela 16 - Comparações entre tamanhos estimados dos grupos de ligação e valores de distância média entre as marcas estimadas a partir da informação de 200 genótipos recombinantes de população de retrocruzamento.

Populações	Grupo de ligação	Tamanho estimado	Tamanho paramétrico	Desvio	Média estimada	Média paramétrica	Desvio
1	1	159,44	150	9,44	5,49	5	0.49
1	2	157,42	150	7,42	5,42	5	0.42
1	3	159,53	150	9,53	5,50	5	0.50
1	4	154,76	150	4,76	5,28	5	0.28
2	1	143,54	150	-6,46	5,20	5	0.20
2	2	151,53	150	1,53	5,50	5	0.50
2	3	151,93	150	1,93	5,52	5	0.52
2	4	168,6	150	18,6	6,15	5	1.15
3	1	153,45	150	3,45	5,58	5	0.58
3	2	167,26	150	17,26	6,11	5	1.11
3	3	146,76	150	-3,24	5,32	5	0.32
3	4	153,63	150	3,63	5,58	5	0.58
4	1	140,54	150	-9,46	5,09	5	0.09
4	2	159,19	150	9,19	5,82	5	0.82
4	3	144,44	150	-5,56	5,23	5	0.23
4	4	143,51	150	-6,49	5,19	5	0.19
5	1	157,8	150	7,8	5,74	5	0.74
5	2	153,07	150	3,07	5,57	5	0.57
5	3	146,34	150	-3,66	5,30	5	0.30
5	4	147,47	150	-2,53	5,34	5	0.34
6	1	153,07	150	3,07	5,57	5	0.57
6	2	147,85	150	-2,15	5,36	5	0.36
6	3	155,79	150	5,79	5,89	5	0.89
6	4	153,11	150	3,11	5,56	5	0.56
Média dos desvios				2,91			0,51

Os desvios dos valores de recombinação dos grupos de ligação e os desvios em relação às médias das distâncias de cada grupo estão de acordo com os valores observados em outros trabalhos utilizando populações obtidas exclusivamente por recombinação estocástica que também indica uma tendência

no aumento do tamanho dos grupos de ligação estimados (SILVA, 2005). Os valores dos desvios também corroboram a expectativa de que as populações obtidas são propícias para as avaliações seguintes de posicionamento e caracterização de QTLs. Não foram observadas alterações na ordem de nenhum dos marcadores como também não foi detectado nenhum marcador não ligado.

Uma vez obtidas as estimativas de recombinação entre marcadores adjacentes e a validação destes valores considerando a estatística, a ordem mais provável dos marcadores acrescido o loco QTL foi determinada utilizando a metodologia “SARF”, que se baseia na escolha da melhor ordem considerando a menor soma de recombinações adjacentes.

Os resultados de mapeamento indicam que os parâmetros definidos para simulação das populações, número de indivíduos e saturação média do mapa, determinados considerando a natureza prática dos trabalhos de mapeamento, permitiram a obtenção de estimativas de recombinação de robustez suficiente para a recuperação da informação paramétrica.

5.5- Detecção de OTLs

Previamente mapeados, os marcadores podem ser avaliados quanto à sua ligação com a característica de herança oligogênica utilizando as novas probabilidades condicionais de ocorrência entre marcador molecular e locos QTL. Visando melhor caracterizar os resultados do método, apresentado nas tabelas 17 e 18, são mostrados os valores de distância estimados entre o QTL e os marcadores alocados nos grupos de ligação assim como o valor do LOD score para estas estimativas. A partir destes valores, é possível detectar as marcas moleculares de maior importância para a expressão da característica oligogênica, em relação às seis condições consideradas.

As tabelas representam um resumo da informação de pronta utilidade para o melhorista de plantas. Nestas tabelas estão contidas a identificação da natureza do marcador d (dominante) ou c (codominante), os valores da estatística LOD que ponderam a probabilidade de ocorrência da hipótese de ligação contra a hipótese

de segregação independente em relação ao loco QTL e os valores de distância estimados entre o loco QTL e a marca molecular. A distância paramétrica entre o loco controlador da característica e os marcadores do grupo de ligação é apresentada para comparação com os valores estimados. Nas populações avaliadas não foram observados erros na identificação dos locos controladores das características, e, em média, as funções de verossimilhança indicaram 4,3 marcadores candidatos por loco controlador.

Este número é aproximadamente 1/4 do número de marcadores candidatos indicados pela técnica de marca simples (dados não mostrados). Embora de natureza semelhante à técnica de marca simples, a similaridades entre os métodos pode ser avaliado pelos significativos valores de correlação entre os valores de LOD e da estatística F da marca simples. O menor número de marcadores candidatos indicados sistematicamente pelas funções de verossimilhança resulta de seu maior poder de teste desta metodologia em relação à marca simples.

Em relação a outras propriedades do método, destacam-se a não necessidade de informação prévia de ordenamento das marcas, o não confundimento de efeitos de QTLs localizados no mesmo grupo de ligação, o maior poder de teste comparado também ao mapeamento por intervalo e a marca simples.

Após a identificação dos putativos grupos aos quais devem pertencer os locos QTL, estes grupos têm sua ordem novamente questionada pelo método SARF, incluída a informação da característica oligogênica, que, por apresentar padrão binário, pode ser considerada como uma marca adicional no mapa, visando obter a melhor estimativa de posição do QTL ao longo dos grupos de ligação.

Uma vez eleita à ordem de maior parcimônia em função dos valores de distância entre as marcas adjacentes e a informação da característica oligogênica, o loco QTL pode ser posicionado como mostrado na Figura 2. A Figura 2 mostra o mapa genético original em azul e os grupos de ligação acrescidos, estando à posição dos QTLs nos grupos de ligação realçada em verde.

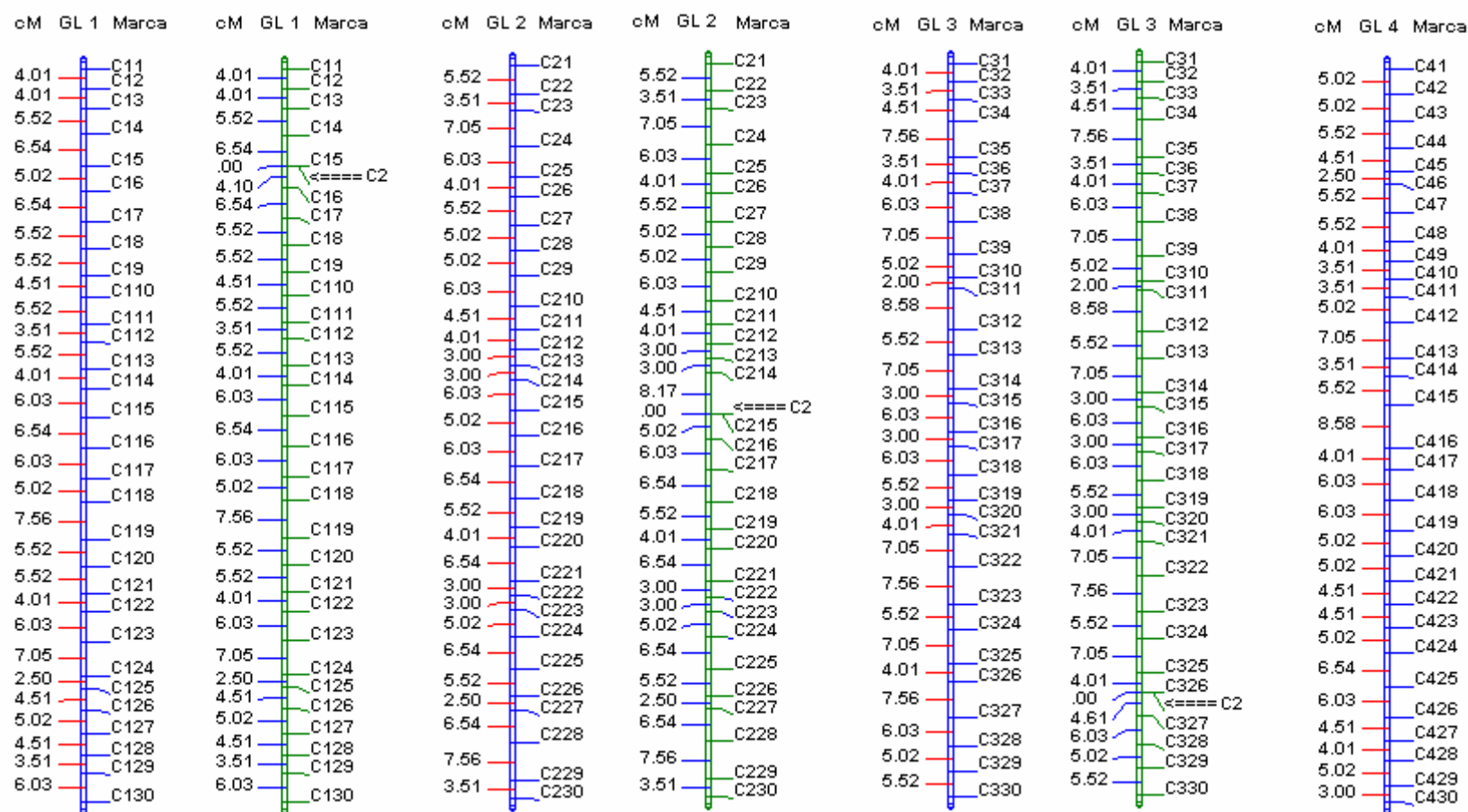


Figura 2 - Mapa genético de população de retrocruzamento estimado a partir da informação de recombinação entre os marcadores provenientes de 200 genótipos. Os grupos de ligação ($LOD = 3$ e $r = 0,30$ e unidade de Kosambi) são os grupos de ligação realçados em verde que foram obtidos pelo método analítico acrescidos da informação de da distância e da posição dos locos controladores da expressão da característica oligogênica de expressão governada por dois locos moleculares identificados como QRL

A incorporação do loco controlador da característica ao mapa de ligação, em geral, resulta em modificações nos valores de recombinação entre marcas adjacentes: quanto maior a distância entre as marcas, menor o número de eventos meióticos analisados. Esta alteração é resultado da inclusão da característica como se fosse uma marca ao mapa, respeitando sua natureza epistática e multifatorial. O aumento na precisão da metodologia é esperado em função do atendimento das pressuposições em relação ao número de locos e da natureza da interação entre alelos de mesmos genes e entre genes.

A maior quantidade de cálculos necessária para a obtenção das estimativas em comparação com métodos baseados em mínimos quadrados, a necessidade de ajuste das funções para os casos não contemplados neste estudo e o conhecimento “a priori” do padrão de herança da característica estudada são as principais limitações deste método.

Tabela 17: Distâncias estimadas dos locos controladores das características oligogênicas dos marcadores moleculares previamente mapeados, utilizando as funções de verossimilhança para cada uma das seis condições avaliadas.

Condição	Marcador	LOD	Distância estimada	Distância paramétrica	Condição	Marcador	LOD	Distância estimada	Distância paramétrica
1 - 9p7	D13	11.22	8.07	10		D213	6.72	4.31	5
	D14	14.08	4.31	5		D214	9.04	0	0
	D15	20	0	0		D114	4.53	5.22	5
	D16	18.19	0	5	2 - 9p7	D15	8.6	6.35	0
	D17	14.08	4.31	10		D16	6.66	0	5
	D18	13.54	4.51	15		D17	4.22	7.22	10
	D19	10	9.93	20		D213	3.12	8.92	5
	D213	15.18	7.76	5		D214	6.58	0	0
	D214	26.62	0	0		D215	5.22	5.77	5
	D215	17.43	5.62	5		D216	3.56	9.23	10
	D216	12.77	9.2	10		D113	3.55	7.88	10
	D217	12.52	9.93	15	2 - 13p3a	D14	7.01	6.58	5
1 - 13p3a	D13	6.22	8.54	10		D15	9.94	0	0
	D14	8.12	7.22	5		D16	9.12	4.47	5
	D15	11.05	0	0		D17	5.22	9.86	10
	D16	10.23	4.12	5		D213	2.89	6.01	5
	D17	6.33	9	10		D214	4.12	0	0
	D18	7.01	10.01	15		D215	2.1	7.83	5
	D213	4	6.23	5		D113	3.94	8.97	10

1 - 13p3b	D214	5.23	0	0	2 - 13p3b	D14	4.91	8.33	5
	D215	3.21	7.44	5		D15	7.78	0	0
	D13	5.15	9.11	10		D16	6.31	6.01	5
	D14	6.12	8.01	5		D17	4.91	8.87	10
	D15	8.99	0	0		D213	2	6.56	5
	D16	7.52	5.23	5		D214	3.23	0	0
	D17	6.12	9.77	10		D215	1.8	7.48	5
	D213	3.21	7.44	5	2 - 15p1	D14	4.22	8.12	5
1 - 15p1	D214	4.44	0	0		D15	9.1	0	0
	D215	3.01	8.43	5		D16	9.47	0	5
	D14	5.54	9.2	5		D17	9.47	0	10
	D15	10.42	0	0		D18	7.32	3.21	15
	D16	10.79	0	5		D19	4.22	9.21	20
	D17	10.79	0	10		D213	5.4	4.12	5
	D18	8.64	2.8	15		D214	7.72	0	0
	D19	5.54	9.2	20	4 - 27p37	D16	11.96	0	5
3 - 9p7	C13	8.79	9.72	10		D17	9.87	2.4	10
	C14	11.27	6.23	5		D18	9.53	2.5	15
	C15	18.8	0	0		D19	8.39	4.71	20
	C16	11.87	6.13	5		D213	9.67	7.35	5
	C211	13.47	7.56	15		D214	16.99	0	0
	C212	15.3	5.62	10		D215	10.54	5.93	5
	C213	21.02	1.4	5		D324	7.04	4.9	5
	C214	26.62	0	0		D325	9.1	0	0
3 - 13p3a	C215	16.63	6.64	5		D326	8.6	0	5
	C13	16.96	7.45	10	5 - 63p1	D327	7.4	7.8	10
	C14	20.63	4.41	5		D15	3.44	0	0
	C15	26.32	0	0		D16	3.29	0	5
	C16	20.62	4.41	5		D214	3.25	1.26	0
	C17	15.33	9	10		D215	4.87	8.88	15
	C18	15.74	8.58	15		D325	3.9	0	5
	C213	6.63	0	5	5 - 27p37	D13	5.93	7.49	10
	C214	6.38	0	0		D14	7.16	5.01	5
3 - 13p3b	C215	6.51	0	5		D15	11.81	0	0
	C216	4.3	3.51	10		D16	10.73	1.22	5
	C19	3.15	9.31	30		D17	8.64	2.77	10
	C111	3.43	6.94	20		D18	8.3	5.68	15
	C112	3.53	6.84	15		D19	7.16	8.01	20
	C113	3.53	6.84	10		D213	8.44	4.35	5
	C114	4.65	3.2	5		D214	15.76	0	0
	C115	6.77	0	0		D215	9.31	6.77	5
3 - 13p3b	C116	6.39	0	5	6 - 63p1	D324	8	6.1	5
	C117	4.65	3.2	10		D325	10.2	0	0
	C118	4.54	3.41	15		D326	8.5	0	5
	C119	3.53	6.84	20		D327	7.2	8.8	10
	C213	21.6	6.33	5		C15	3.42	0	0
	C214	31.95	0	0		C16	3.39	0	5
	C215	28.38	1.6	5		C212	3.22	0	10
	C216	22.08	6.13	10		C213	3.19	0	5

	C217	19.49	8.07	15		C214	3.05	0	0
3 - 15p1	C14	5.82	3.3	5		C324	3.1	0	0
	C15	7.85	0	0		C325	3.08	0	5
	C16	5.69	3.41	5	6 - 27p37	C14	6.52	7.07	5
	C17	4.12	7.45	10		C15	11.14	0	0
	C211	5.22	4.61	15		C16	5.69	9.72	5
	C212	3.92	9.72	10		C211	8.29	6.64	15
	C213	6.85	0	5		C212	9.79	4.01	10
	C214	6.22	0	0		C213	11.71	2	5
						C214	15.71	0	0
						C215	11.14	5.22	5
4 - 63p1	D12	6.1	9.62	15		C216	10.23	6.94	10
	D15	3.2	0	0		C217	9.24	7.25	15
	D16	3.05	0	5		C218	8.08	8.79	20
	D214	3.01	0	0		C323	8.2	0	5
	D215	4.28	6.54	5		C324	8.8	0	0
	D325	3.8	0	0		C325	9.4	0	5
	D326	3.01	0	5		C326	7.8	0	10
4 - 27p37	D13	7.16	7.15	10					
	D14	8.39	4.71	5					
	D15	13.04	0	0					

5.6 Segregação ambígua

Um dos pontos mais importantes para o estudo das características de distribuição discreta é o teste de hipótese genética que permite com uma margem de erro concluir a respeito do padrão predominante de segregação: se governado por um, dois ou mais genes e o tipo de interação predominante entre genes e alelos (SCHUSTER & CRUZ, 2004, LYNCH & WALSH, 1998, LIU, 1997). Neste contexto, denomina-se segregação ambígua a impossibilidade de discriminação de duas ou mais hipóteses genéticas. A probabilidade de ocorrência de segregação ambígua entre duas ou mais hipóteses genéticas distintas é inversamente proporcional ao número de indivíduos avaliados e à dissimilaridade dos padrões de segregação comparados.

A não rejeição de uma hipótese genética que não representa o verdadeiro mecanismo de herança de uma característica pode resultar em erros grosseiros e conclusões equivocadas. Visando caracterizar uma limitação do método proposto e a importância de experimentos de segregação para o estudo de características oligogênicas, as hipóteses genéticas utilizadas para obtenção das funções de verossimilhança foram consideradas par a par segundo o número de indivíduos

necessários para a sua discriminação utilizando o método proposto por BARTLEY (1961), descrito por LIU (1998) (Tabela 18).

Tabela 18: Número de indivíduos necessários para discriminação das hipóteses genéticas produzidas pela segregação de um, dois e três genes, com efeito, epistático em populações F2.

	1p1	3p1	15p1	9p7	13p3a	63p1	27p37
1p1		54	12	974	32	7	619
3p1			50	95	666	22	112
15p1				15	99	235	17
9p7					49	9	15146
13p3a						35	55
63p1							9
27p37							

A estatística de Bartley estima o número mínimo de observações que devem ser avaliadas para separar par a par as distribuições de probabilidade de χ^2 relacionadas com cada uma das hipóteses genéticas. Estas distribuições de probabilidade são tão semelhantes quanto maior for a similaridade das proporções genotípicas comparadas. A avaliação da Tabela 15 mostra que algumas hipóteses são extremamente difíceis de ser discriminadas de outras. A separação categórica das hipóteses genéticas 9:7 de 27:37, por exemplo, necessita da análise de mais de 15.000 observações.

Visando facilitar a interpretação dos dados, foi construído um dendrograma utilizando o inverso do número de indivíduos estimados pela estatística de Bartley como uma medida de dissimilaridade. Tal ferramenta gráfica facilita a visualização da relação entre as hipóteses e subsidia o planejamento de cruzamentos tendo em vista principalmente a separação das hipóteses governadas por diferentes números de genes (Figura 2).

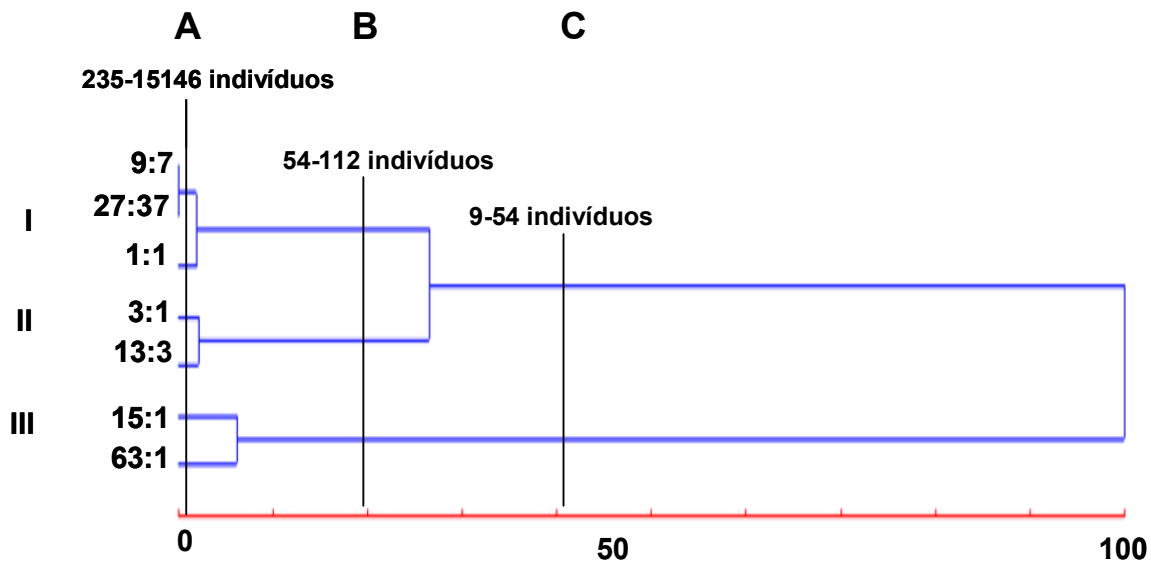


Figura 2- Dendrograma construído pela técnica de UPGMA a partir de matriz de distância estimadas através do inverso do número de indivíduos necessários para separar duas hipóteses genéticas estimadas pela estatística de BARTLEY (1961) citado por LIU (2004).

Uma análise visual do dendrograma permite a discriminação dos pontos de fusão de grupos visando caracterizar as altas mudanças de nível como delimitadoras do número mínimo de indivíduos necessários para a discriminação das hipóteses genéticas. Desta forma, as fusões entre os grupos I, II e III serão interpretadas considerando as linhas transversais traçadas nos pontos de maior mudança de nível, representados por A, B e C (Figura 2).

A linha C representa o ponto de maior mudança de nível que se caracteriza pela discriminação das hipóteses 15:1 e 63:1 das restantes. O número de indivíduos considerados nesta linha (9-54 indivíduos) não fornece subsídios para concluir, nem mesmo, em relação ao número de genes ou quanto ao mecanismo predominante da herança. Por outro lado, um aumento indiscriminado no tamanho das populações pode se mostrar frustrante tendo em vista que a separação de algumas hipóteses pode exigir a avaliação de populações de tamanho bem superior a 100 indivíduos. Além do par 9:7 e 27:37, as hipóteses 1:1 e 9:7, 1:1 e 27:37, 3:1 e 13:3, 3:1 e 27:37, 15:1 e 63:1 dependem da análise de mais de 100

indivíduos para ser discriminadas.

Desconsiderando-se hipóteses genéticas semelhantes, observa-se que a utilização de 50-100 indivíduos em ensaios de segregação piloto permite separação das hipóteses menos similares. Limitações de natureza prática contribuem para que o número de 50 indivíduos esteja próximo do limite da quantidade avaliada em ensaios de segregação de diversas culturas e que alternativas devem ser consideradas visando a um aumento na confiabilidade dos testes de segregação.

A não rejeição de uma hipótese de segregação pelo teste de qui-quadrado pode ser obtida a partir de um número relativamente pequeno de observações. No entanto, a caracterização da herança e a opção pela hipótese genética mais provável devem ser feitas a partir da análise de um conjunto de poucas famílias ou em função de milhares de indivíduos.

O elevado número de indivíduos necessários para a discriminação entre algumas hipóteses pode ser considerado injustificável e de condução praticamente impossível na maioria das situações. Desta forma, a análise simultânea de uma rede de cruzamentos planejados visando avaliar a segregação dos fatores responsáveis pela herança da característica em famílias aparentadas por meio da avaliação de genitores interconectados parece ser a estratégia mais racional para a determinação da hipótese genética mais adequada.

Nesta estratégia, a informação da segregação de uma família é utilizada em complementaridade à de outras, caracterizando assim a natureza da resposta. A Figura 3 é uma representação de uma rede de cruzamentos que objetiva, a partir da análise de duas gerações, validar a herança de características oligogênicas, explorando cruzamentos-teste e autofecundações. A identificação de indivíduos completamente recessivos é questão primordial nesta rede de cruzamentos. A condição homozigota e recessiva destes indivíduos pode ser verificada por uma única geração por autofecundação. A avaliação de uma progênie de autofecundação composta por apenas 6 indivíduos permite concluir em relação ao caráter homozigoto de um loco com probabilidade superior a 99%.

O tempo e a mão de obra necessária para o desenvolvimento de uma rede

de cruzamentos em espécies perenes faz este planejamento ainda mais importante. Neste esquema, foram exploradas principalmente cruzamentos- teste e autofecundações. No entanto, outras configurações também podiam ter sido avaliadas, como famílias de meios-irmãos.

A simples utilização do critério da parcimônia para definição de uma hipótese genética, em que a opção pelo mais simples reflete uma expectativa do pesquisador de maior probabilidade de acerto, pode ser extremamente traiçoeira, considerando, por exemplo, que uma hipótese genética tão simples quanto à segregação de um único gene com proporção esperada de 1:1 necessite de mais de 500 observações para ser discriminada das hipóteses 9:7 (2 genes) e 27:37 (3 genes).

Visando elucidar o efeito da consideração de uma hipótese equivocada na utilização do método, foram realizadas avaliações utilizando populações simuladas utilizadas para a detecção dos QTLs e recuperação da informação genômica.

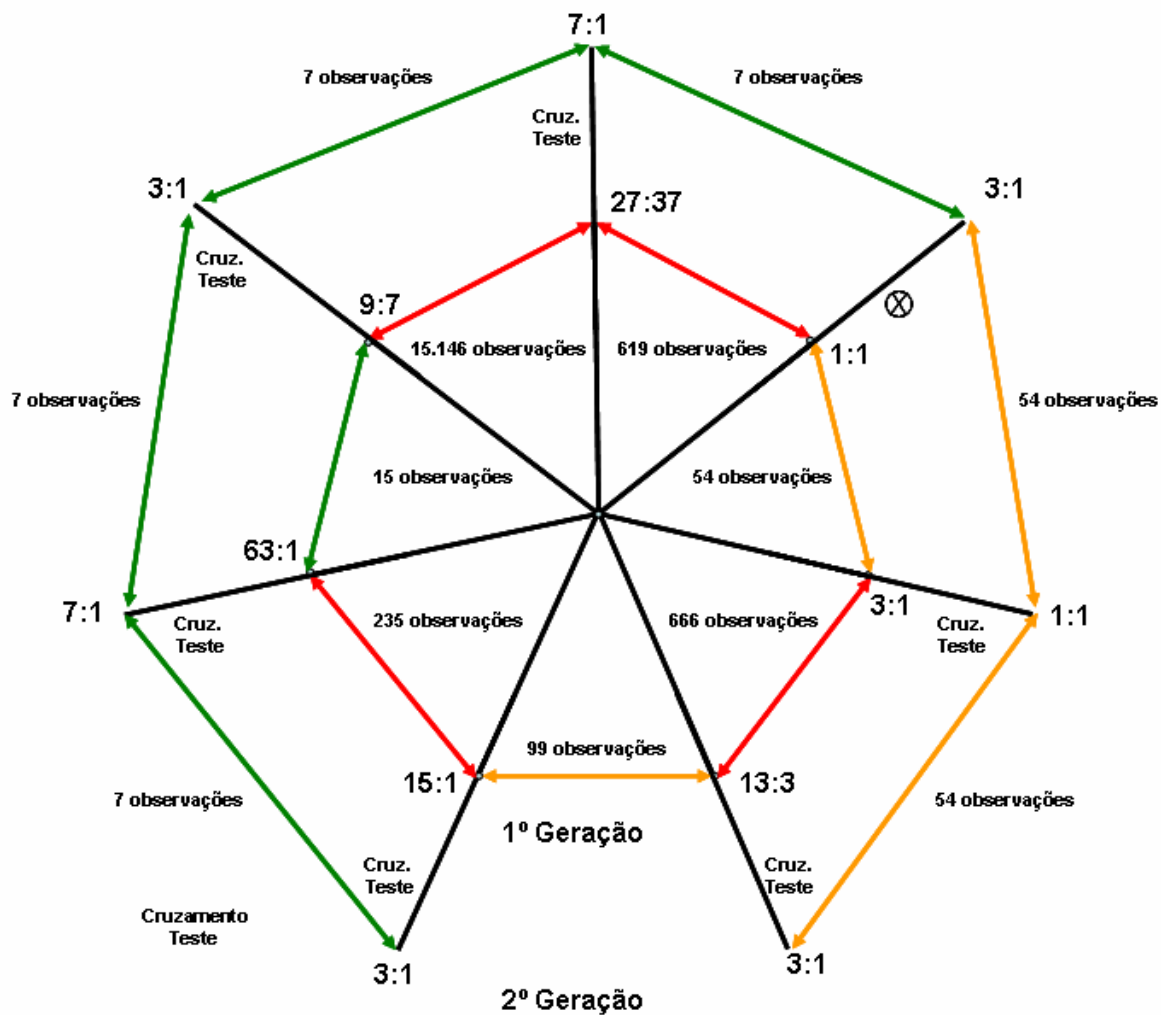


Figura 3 – Representação esquemática do planejamento experimental de uma rede de cruzamentos para elucidação do padrão de herança de características oligogênicas governadas por um, dois e três genes. O heptágono interno representa a 1ª geração e o externo a 2ª geração, sendo que o número de observações necessárias para discriminar as hipóteses genéticas está representado por setas duplas nas cores: verde (7 observações), alaranjada (54-99 observações) e vermelho (acima de 100 observações)

5.7 - Erro na declaração da hipótese genética

Dois tipos de erros estão associados a um teste de hipótese. O erro tipo I associado à probabilidade de não rejeitar uma hipótese quando esta é falsa, denominada por α , e o erro tipo II denominado por β , definido como a probabilidade de não rejeitar a hipótese nula quando esta é falsa (LIU, 1997). Neste contexto, o poder do teste pode ser definido como a probabilidade de rejeitar a hipótese nula quando a hipótese alternativa é verdadeira. Visando quantificar o efeito da adoção de uma hipótese genética equivocada para estudo, foram simuladas 100 populações F2 constituídas por 200 indivíduos e 5 características governadas por 2 genes de interação epistática que produzem as proporções 9:7, 13:3, 15:1 e governadas por 3 genes cujas interações epistáticas produzem as proporções 63:1 e 27:37.

O sucesso na detecção do loco QTL foi avaliado nas populações F2, considerando simultaneamente o estabelecimento “a priori” da hipótese correta e de hipóteses equivocadas em relação à herança da característica oligogênica (Tabela 19). Na diagonal da tabela encontram-se os resultados relativos à análise com a escolha apropriada da hipótese genética e pode ser interpretada como a acurácia do método frente às diferentes proporções epistáticas.

A proporção epistática que apresentou a menor acurácia foi proporção 63:1. Recomenda-se que populações constituídas de número de indivíduos bastante superior a 200 sejam avaliadas para caracterizar a herança de características com esta proporção. A baixa frequência de uma das classes dificulta a avaliação dos eventos meióticos representativos da classe mais rara, fazendo com que as estimativas de recombinação e de detecção de QTL percam bastante de sua acurácia.

A rigor foi considerado um valor de LOD_{min} igual a 2 na proporção 63:1, considerando a inadequação do número de indivíduos para análise desta classe. As proporções de detecção mais acuradas foram respectivamente: 9:7, 27:37, 13:3, 15:1 e 63:1.

A interação epistática 13:3 pode apresentar duas formas diferentes dependendo do gene que se referencia (A ou B, Tabela 1). A escolha pela hipótese 13:3 equivocada impossibilita a detecção de um segundo ou terceiro loco que eventualmente esteja controlando a expressão da característica. A escolha equivocada por interações epistáticas governadas por diferentes números de genes também deve fazer com que fatores importantes sejam excluídos da análise. Em geral, observa-se que a suposição de uma hipótese equivocada resulta para o mapeamento em uma maior frequência de erros tipo I do que erros tipo II.

Além da acurácia na detecção do QTL, uma das vantagens de optar pela utilização do M.M.C.O. é o de não interferir na interpretação de uma herança que sabidamente é governada por poucos fatores de maior efeito. A avaliação, utilizando métodos quantitativos tradicionais de detecção de QTLs, dificulta a caracterização da herança oligogênica e pode levar o pesquisador a supor uma herança em relação a seu caráter. Assim, entende-se que o aumento no poder do teste associado a uma facilidade relativa de discriminação das proporções faz da metodologia baseada em funções de verossimilhança uma alternativa bastante atrativa para a caracterização da herança das características oligogênicas.

Tabela 19 – Proporção em 100 populações da rejeição da hipótese H_0 de segregação independente entre os locos responsáveis pela expressão e a característica oligogênica avaliada e de ocorrência de erros tipo I e tipo II

Loco		Interações epistáticas														
		9p7			13p3			15p1			63p1**			27p37		
A (C15)		Ho _(QRL C15)	tipo I	tipo II	Ho _(QRL C15)	Tipo I	tipo II	Ho _(QRL C15)	tipo I	tipo II	Ho _(QRL C15)	tipo I	tipo II	Ho _(QRL C15)	tipo I	tipo II
	9p7	100	0	0	1	94	5	2	94	4	0	100	0	99	1	0
	13p3	64	36	0	94	6	0	96	3	1	9	91	0	91	9	0
	15p1	0	100	0	0	100	0	90	10	0	68	14	18	2	98	0
	63p1	0	100	0	9	91	0	68	14	18	73	27	0	0	100	0
	27p37	99	1	0	91	9	0	2	98	0	0	100	0	98	2	0
		9p7			13p3			15p1			63p1**			27p37		
B (C214)		Ho _(QRL C15)	tipo I	tipo II	Ho _(QRL C15)	Tipo I	tipo II	Ho _(QRL C15)	tipo I	tipo II	Ho _(QRL C15)	tipo I	tipo II	Ho _(QRL C15)	tipo I	tipo II
	9p7	100	0	0	64	36	0	2	94	40	0	100	0	99	1	0
	13p3	69	31	0	94	6	0	0	99	1	0	100	0	48	52	0
	15p1	0	100	0	0	99	1	90	10	0	68	14	18	2	98	0
	63p1	0	100	0	0	100	0	68	14	18	73	27	0	0	100	0
	27p37	99	1	0	48	52	0	2	98	0	0	100	0	98	2	0
		9p7			13p3			15p1			63p1**			27p37		
C (C326)		Ho _(QRL C15)	tipo I	tipo II	Ho _(QRL C15)	Tipo I	tipo II	Ho _(QRL C15)	tipo I	Tipo II	Ho _(QRL C15)	tipo I	tipo II	Ho _(QRL C15)	tipo I	tipo II
	9p7	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	13p3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	15p1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	63p1	0	100	0	0	100	0	0	100	0	73	27	0	0	100	0
	27p37	0	100	0	0	100	0	0	100	0	0	100	0	98	2	0

** Pelo fato de uma das classes ser de ocorrência rara é necessário à análise de populações de tamanho superior a 200 indivíduos para obtenção de valores de LOD superior a 3. O valor de LODmin igual a 2 foi considerado neste caso para a confecção desta tabela.

5.8. Estudo de herança e detecção de OTLs

Tradicionalmente considera-se que a caracterização precisa da herança deve preceder os trabalhos de mapeamento e detecção de QTLs, uma vez que o planejamento de cruzamentos, o estabelecimento de populações experimentais e a escolha apropriada dos métodos de análise dependem do conhecimento a respeito das propriedades da herança.

Por outro lado, entende-se que a informação de marcadores moleculares pode em muito contribuir para a detecção de fatores de maior efeito relacionado com a expressão das características previamente às análises de mapeamento e detecção de QTLs, subsidiando desta forma todas as pressuposições que o pesquisador deve assegurar em suas análise, inclusive a da escolha correta da hipótese genética. As análises realizadas com técnicas de detecção de QTLs de natureza simples, como a técnica de marca simples, por exemplo, mostram que tais estratégias são indicadas para a detecção de QTLs em análises iniciais e exploratórias em que o rigor da caracterização do QTL não é essencial, e que somente em uma segunda etapa, técnicas mais acuradas como as funções de verossimilhança seriam utilizadas para detecção de locos de maior importância para a expressão de características de natureza oligogênica.

Desta forma, um estudo de uma herança poderia ser descrito em etapas como se segue:

1- Estabelecimento de populações contrastantes para a característica oligogênica em questão interconectadas por genitores em comum e estabelecimento de populações autofecundadas a partir de genitores que participam de maior número de cruzamentos.

2- Análise exploratória visando à caracterização precisa do padrão de herança valendo-se das seguintes ferramentas:

2.1- Análise conjunta de populações interconectadas e progênes autofecundadas:

2.3 - Análise utilizando marcadores moleculares: A utilização de marcadores moleculares espalhados por todo o genoma (1 marcador a cada 30

cM) associado à análise exploratória de marca simples favorece a identificação dos fatores mais importantes.

2.4 – Pressuposição de uma hipótese de herança e utilização de técnicas de maior acurácia para a detecção e caracterização do QTL, possivelmente associado à avaliação de um maior número de indivíduos de progênes selecionadas.

6. CONCLUSÕES

Além do método apresentado neste trabalho, outros também poderiam ser utilizados para a detecção de características oligogênicas. Algumas características contribuem para a investigação de padrões oligogênicos utilizando a estratégia proposta visto que:

- 1- Permite caracterizar o padrão de herança de uma característica evitando o erro de interpretar de maneira contínua uma distribuição que sabidamente possui classes;
- 2- Possui maior poder de teste que métodos exploratórios, tais como o método da marca simples;
- 3- Não necessita da informação prévia de mapeamento;
- 4- Não é influenciada pela ocorrência de dois QTLs no mesmo grupo de ligação;
- 5- Pode ser adequada para quaisquer tipos de populações e padrões de segregação; e
- 6- Agrega as vantagens dos estimadores de máxima verossimilhança.

7. REFERÊNCIAS BIBLIOGRÁFICAS

- ALFENAS A.C., ZAUZA, E.A.V., MAFIA R.G., ASSIS T.F. **Clonagem e doenças do Eucalipto**. Viçosa, MG. Editora UFV, 2004. 442p.
- ALZATE-MARIN A.L., CERVIGNI G.D.L., MOREIRA M.A., BARROS E.G. Seleção assistida por marcadores moleculares visando ao desenvolvimento de plantas resistentes a doenças, com ênfase em feijoeiro e soja. **Fitopatologia Brasileira**, v.30, n.4, p.333-342, 2005.
- BADU R., SUDHA K.N., PRASANNA B.M., GUPTA H.S. Integrating marker assisted selection in crop breeding – Prospects and challenges. **Current Science**, v.87, n.5, p.607-619, 2004.
- BOREM, A. **Melhoramento de plantas**. Editora UFV. 3ª Edição. Viçosa, MG. UFV, 2001. 500 p.
- BRONDANI, R.P.V., BRONDANI C., TARCHINI R., GRATTAPAGLIA D. Development, characterization and mapping of microsatellite markers in *Eucalyptus grandis* and *E. urophylla*. **Theoretical and Applied Genetics** v.97: p.816-827, 1998.
- BUNYAMIN T., THOMAS E.M., PAULS K.P. Genetic mapping of agronomic traits in common bean. **Crop Science**, v.42, p.544-556, 2002.
- CERVERA M.T., STORME V., IVENS B., GUSMÃO J., LIU B.H., HOSTYN V., SLYCKEN J.V., MONTAGU M.V., BOERJAN W. Dense Genetic Linkage Maps of Three Populus Species (*Populus deltoides*, *P. nigra* and *P. trichocarpa*) Based on AFLP and Microsatellite Markers. **Genetics** v.158 p.787-809, 2001.
- CHARLESWORTH, B., SNIEGOWSKI, P., STEPHAN, W. The evolutionary dynamics of repetitive DNA in eukaryotes. **Nature**, v.371, p.215 – 220, 1994.
- CRUZ, C. D. **Programa para análises de dados moleculares e quantitativos – GQMOL**. Viçosa: UFV, 2007.(Software em desenvolvimento).
- CRUZ C.D., REGAZZI A.J., CARNEIRO P.C.S. **Modelos Biométricos Aplicados ao Melhoramento Genético - Volume 1**, Viçosa , MG. Editora UFV, 2004. 480p.
- CRUZ C.D., REGAZZI A.J., CARNEIRO P.C.S., **Modelos Biométricos Aplicados ao Melhoramento Genético - Volume 2**, Viçosa , MG. Editora UFV, 2003. 585p.
- DEKKERS J.C.M., HOSPITAL F. The use of molecular genetics in the improvement of agricultural populations. **Nature reviews**, v.3, p.22-32, 2001.

- FALCONER, D.S. **Introdução à genética quantitativa**. Trad. De Martinho de Almeida e Silva e José Carlos Silva. Viçosa, MG: UFV, 1987. 279p.
- HARUSHIMA Y., YANO M., SHOMURA A., SATO M., SHIMANO T., KUBOKI, Y., YAMAMOTO T., LIN S.Y., BALTAZAR A.A., PARCO A., KAJIYA H., HUANG N., YAMAMOTO K., NAGAMURA Y., KURATA N., KHUSH G.S., SASAKI T.A. high-density rice genetic linkage map with 2275 markers using a single F2 population. **Genetics**, v.148, p.479-494, 1998.
- JANSEN, R.C. Interval mapping of multiple quantitative trait loci. **Genetics**. v.135, p.205 -211, 1993.
- JUNGHANS D. T., ALFENAS A. C., BROMMONSHENKEL S. H., ODA S., MELLO E. J., GRATTAPAGLIA D. Resistance to rust (*Puccinia psidii* Winter) in Eucalyptus mode of inheritance and mapping of a major gene with RAPD markers. **Theoretical and Applied Genetics**. v.108, 175-180. 2003
- LANDER E. S., BOTSTEIN D., Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. **Genetics** v.121, n.1, p185-199, 1989.
- LI J., **Development of multiple interval mapping for mapping QTL in ordinal traits**. Raleigh, North Carolina, North Carolina State University, 2004. 124p.
- LIU B.H., **Statistical Genomics: linkage mapping and QTL analysis**. Boca raton, Florida, USA: CRC Press, 1997. 610p.
- LYNCH, M.; WALSH, B. **Genetics and analysis of quantitative traits**. Sunderland, USA: Sinauer Associates, Inc.Ed., 1998., 978p.
- MACKAY T. F. C., Genetic architecture of quantitative traits. **Annual Review of Genetics**, v.35, p.303-339; 2001.
- MEKSEM, K., NJITI, V.N., BANZ, W.J., IQBAL, M.J., KASSEM, M.M., HYTEN, D.L., YUANG, J., WINTERS, T.A., LIGHTFOOT, D.A. Genomic regions that underlie soybean seed isoflavone content. **Journal of Biomedicine and Biotechnology**, v.1, n.1, p.38-44, 2001.
- MISSIAGIA A. A., **Mapeamento Genético de QTL para qualidade da madeira e florescimento precoce e estudos de expressão gênica alelo específica em *Eucalyptus* spp.** Piracicaba, 2005. 236p. Dissertação (Doutorado), Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo.
- MURO ABAD, J.I., ROCHA, R.B., CRUZ, C.D., ARAÚJO, E.F. Obtainment of Eucalyptus spp. hybrids aided by molecular markers - SSR analysis. **Scientia Forestalis**. , v.67, p.53 - 63, 2005a.

- MURO ABAD, J.I., ROCHA, R. B., CRUZ, C. D., ARAÚJO, E. F. Crosses recommendation method for obtaining *Eucalyptus* spp. hybrids assisted by molecular markers. **Crop Breeding Applied Biotechnology**, v.5, p.71 - 73, 2005b.
- NOVAES E. **Mapeamento de QTLs para qualidade da madeira em *Eucalyptus grandis* x *E. urophylla* e ancoragem de clones BAC no mapa genético.** Viçosa, 2006. 171p. Dissertação (Mestrado), Universidade Federal de Viçosa.
- RAO, S., LI X., Strategies for genetic mapping of quantitative traits. **Genetica**, v.109, p.183–197, 2000.
- ROCHA R. B., MURO ABAD, J. I., ELEOTÉRIO I. P., ARAÚJO E. F. Fingerprint and genetic diversity analysis of *Eucalyptus* spp. genotypes using RAPD and SSR markers, **Scientia Forestalis** v.62, p.24-31, 2002.
- ROCHA, R.B., CRUZ, C.D., BARROS, W. S., FERREIRA, F.M., ARAÚJO, E.F. Comparisons of segregating populations for genetic mapping. **Crop breeding and applied biotechnology**., v.4, p.408 - 415, 2004.
- SILVA, L.C. **Simulação do tamanho da população e da saturação do genoma para mapeamento genético de RILs.** Viçosa, 2005. 120p. Dissertação (Mestrado), Universidade Federal de Viçosa.
- SCHUSTER I., CRUZ C. D. **Estatística genômica aplicada a populações derivadas de cruzamentos.** Viçosa , MG. Editora UFV, 2004. 585p.
- VARSHNEY R.K., GRANER A., SORRELLS M.E., Genic microsatellite markers in plants: features and applications. **Trends in Biotechnology**, v.23, n.1, p.48-55, 2005.
- VIJAYALAKSHMI S., YADAV K., KUSHWAHA C., SARODE S.B., SRIVASTAVA C.P., CHAND R., SINGH B.D. Identification of RAPD markers linked to the rust (*Uromyces fabae*) resistance gene in pea (*Pisum sativum*). **Euphytica**, v.144, p.265-274, 2005.
- VISSCHER, P. M., C. S. HALEY, AND S. A. KNOTT, Mapping QTLs for binary traits in backcross and F2 populations. **Genetical Research**, v.68, p.55–63, 1996.
- YOUNG N. D. A cautiously optimistic vision for marker-assisted breeding. **Molecular breeding**. v.5, p.505-510, 1999.
- ZENG, Z.B. Precision mapping of quantitative trait loci. **Genetics**, v.136, p.1457-14, 1994.

WEIR B.S., **Genetic data analysis II**. Sinauer Associates, Sunderland, 445p. 1996.

XU, S., W. R. ATCHLEY, Mapping quantitative trait loci for complex binary diseases using line crosses. **Genetics**, v.143, p.1417–1424, 1996.

YI, N., S. XU, V. GEORGE, AND D. B. ALLISON Mapping multiple quantitative trait loci for ordinal trait. **Behavior Genetics**, v.34, p.3–15, 2004.

YU K., PARK S. J., POYSA V. Marker assisted selection of common beans for resistance to common bacterial blight: efficacy and economics. **Plant breeding**, v.199, p.411-415, 2000.