

ELIEL ALVES FERREIRA

EFICIÊNCIA DO TAMANHO DE POPULAÇÕES F₂ NA ANÁLISE DE QTL

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 Science*.

VIÇOSA
MINAS GERAIS – BRASIL
2012

**Ficha catalográfica preparada pela Seção de Catalogação e
Classificação da Biblioteca Central da UFV**

T

F413e
2012

Ferreira, Eliel Alves, 1982-

Eficiência do tamanho de populações F_2 na análise de
QTL / Eliel Alves Ferreira. – Viçosa, MG, 2012.
viii, 118f. : il. ; 29cm.

Inclui apêndices.

Orientador: Cosme Damião Cruz.

Tese (doutorado) - Universidade Federal de Viçosa.

Inclui bibliografia.

1. Locos de caracteres quantitativos - Métodos estatísticos.
2. Genética quantitativa. I. Universidade Federal de Viçosa.
II. Título.

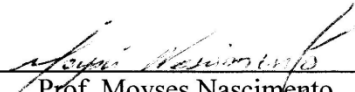
CDD 22. ed. 572.8633

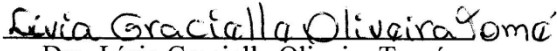
ELIEL ALVES FERREIRA

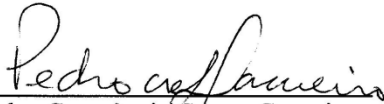
EFICIÊNCIA DO TAMANHO DE POPULAÇÕES F₂ NA ANÁLISE DE QTL

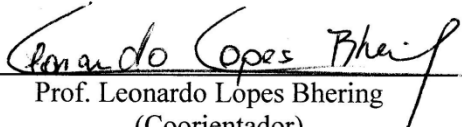
Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Genética e Melhoramento, para obtenção do título de *Doctor Science*.

APROVADO: 27 de fevereiro de 2012.


Prof. Moyses Nascimento


Dra. Livia Gracielle Oliveira Tomé


Prof. Pedro Crescêncio Souza Carneiro
(Coorientador)


Prof. Leonardo Lopes Bhering
(Coorientador)


Prof. Cosme Damião Cruz
(Orientador)

AGRADECIMENTOS

Primeiramente, agradeço a Deus pela vida, saúde e por sempre estar ao meu lado em todos os momentos.

Aos meus pais Mauro e Neusa (*in memoriam*), irmãos e amigos Daniel e Elizanéia pela amizade, apoio e compreensão que foram muito importantes para a realização deste trabalho.

Aos meus cunhados Luciana e Ricardo e à Nilva, pela amizade, apoio e compreensão que foram importantes para a realização deste trabalho.

Ao prof. Cosme Damião Cruz pela amizade e orientação, indispensáveis para a conclusão desta tese.

Aos coorientadores prof. Pedro Crescêncio Souza Carneiro e prof. Leonardo Lopes Bhering, pela amizade e ensinamentos.

À Universidade Federal de Viçosa - UFV e à Pós-Graduação em Genética e Melhoramento pela oportunidade a mim concedida para realização do curso de Doutorado.

Aos professores da Pós-Graduação em Genética e Melhoramento pela contribuição em minha formação.

Aos amigos do laboratório de Bioinformática, em especial ao Caio, Daniele, Felipe, Gilson, Lívia e Rafael, pela amizade, companheirismo, colaboração e auxílio durante o curso.

Aos amigos e colegas da Pós-Graduação, pelo companheirismo e amizade durante o curso.

Às secretárias da Pós-Graduação em Genética e Melhoramento, Rita e Edina, pelo apoio, dedicação, atenção e amizade.

À Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), pela bolsa de Doutorado a mim concedida.

BIOGRAFIA

Eliel Alves Ferreira, filho de Mauro da Fonseca Ferreira e Neusa Olinda Mariano Ferreira, nasceu em Monte Santo de Minas, Minas Gerais, em 17 de abril de 1982.

No ano de 1998, iniciou seus estudos no curso Técnico em Agropecuária na Escola Agrotécnica Federal de Muzambinho, Minas Gerais, concluído no ano de 2000.

Em fevereiro de 2001, em Passos, Minas Gerais, ingressou no curso de Agronomia da Universidade do Estado de Minas Gerais, onde se graduou em dezembro de 2005, obtendo o título de Engenheiro Agrônomo.

Em março de 2006, ingressou no Programa de Pós-Graduação em Agricultura Tropical e Subtropical, área de concentração em Genética, Melhoramento Vegetal e Biotecnologia, no Instituto Agronômico – IAC, em Campinas, São Paulo, onde obteve, em Janeiro de 2008, o Título de Mestre em Ciências.

Em março de 2008, ingressou no Programa de Pós-Graduação em Genética e Melhoramento na Universidade Federal de Viçosa, Minas Gerais, onde obteve, em fevereiro de 2012, o Título de Doutor em Ciências.

SUMÁRIO

RESUMO	vii
ABSTRACT	viii
1. INTRODUÇÃO GERAL	1
2. REVISÃO DE LITERATURA	3
2.1. Mapeamento Genético	3
2.2. Populações de mapeamento.....	5
2.2.1. Metodologias para detecção de QTL.....	6
2.2.1.1. Método da marca simples.....	6
2.2.1.2. Método de estudo de QTL por análise de intervalo	9
2.4. Tamanho de população para detecção de QTL.....	11
2.5. Detecção de QTLs fantasmas	13
3. REFERÊNCIAS BIBLIOGRÁFICAS	15
CAPÍTULO 1	22
TAMANHO DE POPULAÇÃO IDEAL PARA DETECÇÃO DE QTL EM POPULAÇÕES F_2	22
1. INTRODUÇÃO	25
2. MATERIAL E MÉTODOS	27
2.1. Simulação	27
2.1.1. Simulação dos dados	27
2.1.2. Simulação dos genitores.....	27
2.1.3. Procedimentos para simulação dos indivíduos F_2	28
2.1.4. Simulação da característica quantitativa (QTL).....	29
2.2. Mapeamento.....	30
2.2.1. Análise de segregação	30
2.2.2. Estimação da porcentagem de recombinação	31
2.2.3. Determinação dos grupos de ligação	32
2.2.4. Ordenamento das marcar no grupo de ligação.....	33
2.3. Análise de detecção de QTLs.....	33
2.3.1. Método da Marca simples	34
2.3.2. Método de intervalo simples	34
2.3.3. Método de intervalo composto	35
2.4. Eficiência do tamanho populacional sobre o mapeamento e obtenção de estimativas de efeitos de QTLs.....	36

2.5. Eficiência do tamanho populacional na detecção de QTL	37
3. RESULTADO E DISCUSSÃO	38
3.1. Teste de segregação	38
3.2. Recuperação dos grupos de ligação	41
3.3. Distância entre marcas adjacentes	42
3.4. Detecção de QTL	43
3.4.1. Detecção de QTL por marca simples baseada em regressão linear	43
3.4.1.1. Detecção de marcadores ligados a QTLs.....	44
3.4.1.2. Quantificação dos efeitos do QTL	46
3.4.1.3. Comparação entre efeitos do QTL estimados e paramétricos	49
3.4.2. Detecção de QTL por intervalo	52
3.4.2.1. Posição do QTL	52
3.4.2.2. Efeitos dos QTLs	60
3.4.2.3. Comparação entre efeitos do QTL estimados e paramétricos	63
4.4.3. Considerações referentes à detecção de QTLs nas populações F_2	67
4. CONCLUSÕES.....	69
5. REFERÊNCIAS BIBLIOGRÁFICAS	70
CAPÍTULO 2	74
EFICIÊNCIA NA DETECÇÃO DE VÁRIOS QTLs EM UM GRUPO DE LIGAÇÃO GÊNICA	74
1. INTRODUÇÃO	77
6. MATERIAL E MÉTODOS	79
2.1. Simulação	79
2.1.1. Simulação dos dados	79
2.1.2. Simulação dos genitores.....	79
2.1.3. Procedimentos para simulação dos indivíduos F_2	81
2.1.5. Simulação da característica quantitativa (QTL).....	81
2.2. Mapeamento.....	83
2.2.1. Análise de segregação	83
2.2.2. Estimação da porcentagem de recombinação	84
2.2.4. Determinação dos grupos de ligação	85
2.2.5. Ordenamento das marcar no grupo de ligação.....	86
2.3. Análise de detecção de QTLs.....	86
2.3.1.1. Método da Marca simples	87

2.3.1.2. Método de intervalo simples	87
2.3.1.3. Método de intervalo composto	88
2.4. Eficiência na obtenção das estimativas dos efeitos e posicionamento dos QTLs	89
3. RESULTADOS E DISCUSSÃO	91
3.1. Análise das marcas moleculares	91
3.2. Detecção de QTL	94
3.2.1. Detecção de QTL por marca simples.....	94
3.2.2. Detecção de QTL por metodologias de intervalo	97
3.3. Considerações referentes à detecção de QTLs fantasmas.....	106
4. CONCLUSÕES.....	108
5. REFERÊNCIAS BIBLIOGRÁFICAS	109
APÊNDICES	112

RESUMO

FERREIRA, Eliel Alves, D.Sc., Universidade Federal de Viçosa, fevereiro de 2012. **Eficiência do tamanho de populações F_2 na análise de QTL.** Orientador: Cosme Damião Cruz. Coorientadores: Pedro Crescêncio Souza Carneiro e Leonardo Lopes Bhering.

Desde os trabalhos realizados por Mendel, a genética vem se desenvolvendo rapidamente, pois têm sido estabelecidas as relações entre cromossomos e genes, associação entre marcas moleculares às características fenotípicas, construção de mapas genéticos, detecção de genes controladores de características quantitativas (QTL) entre outros, com o advento da Genômica. Para os estudos, utilizando marcadores moleculares, principalmente quando se busca detectar QTL, o tamanho da população é de suma importância, pois quanto maior a população mais precisa será a detecção. Outro ponto importante é a prevenção de detecção de QTL fantasma, sendo este um falso positivo. Assim, objetivou-se, com este trabalho, estudar a detecção de QTLs, por meio de dados simulados, em diferentes tamanhos populacionais referentes ao tamanho populacional e à prevenção da detecção de QTLs inexistentes em populações F_2 . Para isto, foram simulados dois cenários. Com o intuito de verificar qual o tamanho de população é mais eficiente para detecção de QTL, foram simulados quatro tamanhos (200, 400, 600 e 1000 indivíduos) com as mesmas características, referentes ao primeiro cenário. Para o segundo, foi simulada uma população de 1000 indivíduos com um genótipo fictício constituído de quatro grupos de ligação com dois QTLs em três grupos, sendo que um não continha QTL, para controle das metodologias de detecção. Os grupos de ligação 1, 2 e 3 continham dois QTLs, contribuindo com 28% da característica, um contribuindo com 28% e outro com 5% e dois contribuindo com 5% da característica, respectivamente. Para a detecção dos QTLs foram utilizadas as metodologias da marca simples, intervalo simples e intervalo composto, quando necessário. Foram observados que quanto maior o tamanho da população mais precisa é a detecção e o posicionamento do QTL; portanto, a de 1000 indivíduos obteve os melhores resultados. Em relação ao segundo cenário, foi verificado que a metodologia de intervalo composto melhora a não detecção de QTL fantasma e que quanto menor for a contribuição dos QTLs maior será a probabilidade de encontrar QTL fantasma.

ABSTRACT

FERREIRA, Eliel Alves, D.Sc., Universidade Federal de Viçosa, February, 2012. **QTLs detection in F₂ population arising of simulation data.** Adviser: Cosme Damião Cruz. Co-advisers: Pedro Crescêncio Souza Carneiro and Leonardo Lopes Bhering.

Since the work done by Mendel, genetics has been developing quickly. Were established relations between chromosomes and genes, the molecular markers association and phenotypic characteristics, production of gene maps, detection of genes controlling quantitative traits (QTL), among others with the advent of genomics. Studies using molecular markers, especially when it seeks to detect QTL, population size is of paramount importance, because the higher the population the more accurate detection. Another important point is to prevent ghost QTL detection, which is a false positive. The objective this work was to study the QTLs detection, using simulated data, referent to population size and prevention of ghost QTLs in F₂ populations. Two scenarios were simulated. To determine what population size is more efficient, four sizes were simulated (200, 400, 600 and 1000 individuals) with the same characteristics, for the first scenario. Were simulated, for the second scenario, a population of 1000 individuals. It was simulated a genotype with four linkage groups containing two QTLs in three groups and in one group does not contain QTL. In linkage groups 1, 2 and 3, contained two QTLs of large effect, one large and one small end and two small effect, respectively. For QTLs detection, it were used the methodologies of simple mark, simple and composite interval. Were observed that the larger the population size, the better was positioning of QTL and therefore that of 1000 individuals achieved the best results. In the second scenario, it was found that the method of composite interval is more effective in preventing ghost QTL and that the smaller the effect of the QTL more probability find ghost QTL.

1. INTRODUÇÃO GERAL

Desde o final do século XIX e início do século XX, vários pesquisadores estudaram os mecanismos da herança, tomando por base o trabalho proposto por Mendel. Esses estudos também estabeleceram as relações entre cromossomos e genes. Com o aprimoramento dos conhecimentos, mais tarde, Morgan percebeu que a proporção da progênie recombinante variava bastante, e propôs que estas variações, de alguma forma, refletiam a distância entre os genes no cromossomo. Em 1913, A.H. Sturtevant, interpretando dados oriundos da segregação de genes ligados, sugeriu o uso da porcentagem de recombinantes como indicador quantitativo da distância linear entre dois genes na construção de mapas genéticos. Desde então, o mapeamento tornou-se, rapidamente, uma poderosa ferramenta para os geneticistas. Estes mapas eram fundamentados em marcadores morfológicos e citológicos, sendo limitado, mas a partir da década de 80, com o advento dos marcadores de DNA, o mapeamento genético tornou-se ilimitado para praticamente todas as espécies.

Após a construção dos mapas genéticos, foram verificadas regiões nas quais encontram-se genes com efeitos genéticos aditivos, dominantes e epistáticos, e estas regiões são denominadas de QTL (*Quantitative Trait Locus*). A sua identificação é realizada por meio de experimentos que envolvem três fontes de variação: estrutura de pedigree, informações de marcadores moleculares e fenótipos.

Normalmente as características de maior importância agrônômica são quantitativas, podendo ser controladas por um ou mais QTL localizados no mesmo ou em cromossomos diferentes, dificultando o processo de seleção. Diante dessas dificuldades houve a necessidade de maior aprimoramento nesta área, surgindo, assim, a Genômica.

Genômica é a denominação dada à ciência que estuda o genoma de forma completa, pela interação de várias áreas tradicionais da genética, como a Genética Mendeliana, a Citogenética, a Genética Molecular, a Genética de Populações e a Genética Quantitativa, incluindo também a ciência da computação e os sistemas automatizados. Os profissionais que tratam da genômica atuam em pelo menos três áreas, quais sejam: na Genética Clássica, em que são estudados os marcadores genéticos e feitas as análises de ligação, o ordenamento dos genes, o mapeamento genético e o mapeamento de QTLs; na Informática Genômica, que investe na formação de banco de dados, em que se realizam as comparações de sequências e se promove a comparação de

dados e automação de maneira geral; nas Análises de sequência de DNA, em que são realizados os sequenciamentos, os alinhamentos e as comparações de sequências (SCHUSTER; CRUZ, 2008).

A Genômica é uma ferramenta importante no melhoramento, pois com esta tecnologia se consegue identificar e posicionar o QTL de interesse, possibilitando o melhoramento assistido por marcadores moleculares. Para que este tenha sucesso é importante que a detecção e o posicionamento do QTL sejam confiáveis, ao contrário pode-se perder tempo e recursos financeiros. A eficácia na identificação e no posicionamento desses depende, diretamente, da saturação do genoma, do tamanho da população e da metodologia aplicada à detecção de QTL.

Em relação ao tamanho das populações para mapeamento, são encontrados trabalhos na literatura, por meio de simulação, que recuperam os mapas de ligação com populações a partir de 100 indivíduos, dependendo da saturação do genoma e da estrutura da população. Apesar deste tamanho recuperar os grupos de ligação, não se sabe se são eficientes para detectar e posicionar QTL, pois o tamanho populacional para detecção de QTL é pouco estudado.

Outro ponto importante a ser considerado é a metodologia de detecção de QTL, principalmente quando existem mais de um QTL por grupo de ligação. A metodologia de intervalo simples detecta e posiciona o QTL de forma eficiente. Entretanto, este método avalia um intervalo de cada vez, sem considerar o restante da informação do genoma, podendo apresentar problemas na detecção quando existirem outros QTLs. Assim, quando múltiplos QTLs estão presentes, um problema comum é a observação de vários falsos positivos, denominados de QTL fantasma (SCHUSTER; CRUZ, 2008; JANGARELLI et al., 2011).

Do exposto, objetivou-se, com este trabalho, o estudo da detecção, localização, efeito e importância de QTL, por meio de dados simulados, em relação a diferentes tamanhos populacionais e a prevenção da detecção de QTLs inexistentes, ou QTLs fantasmas, em populações F_2 .

2. REVISÃO DE LITERATURA

2.1. Mapeamento Genético

No final do século XIX, Hugo de Vries, Carl Correns e Erick von Tschermak-Seysenegg realizaram inúmeras pesquisas a fim de esclarecer os mecanismos básicos da herança, tomando por base o trabalho proposto por Mendel. W. Sutton e T. Boveri, em 1902, estabeleceram a relação entre cromossomos e fatores mendelianos (genes). Já W. Bateson, E.R. Saunders e R.C. Punnet, também em 1902, demonstraram, em ervilha-doce, que as segregações das características cor da flor e formato do pólen não ocorriam de forma independente (LANDER; WEINBERG, 2000).

Em 1910, T.H. Morgan e colaboradores, trabalhando com *Drosophila melanogaster*, observaram distorções da segunda lei de Mendel. Dessa forma, sugeriram que alguns genes estariam situados no mesmo cromossomo e que durante a meiose, ocasionalmente, ocorreriam, entre os homólogos, trocas de segmentos denominados crossing-over ou permuta. Ao observar que as proporções das progênes recombinantes variavam consideravelmente, propuseram que estas variações, de alguma forma, refletiam a distância entre os genes no cromossomo. Em 1913, A.H. Sturtevant, interpretando dados oriundos da segregação de genes ligados, sugeriu o uso da porcentagem de recombinantes como indicador quantitativo da distância linear entre dois genes na construção de mapas genéticos. Os mapas mostravam que a posição dos genes correspondia à sua ordem linear nos cromossomos. Assim, o conceito de localização dos genes em uma ordem linear passou a ser incorporado à teoria cromossômica da herança (GRIFFITHS et al., 1998).

O mapeamento genético tornou-se, rapidamente, uma poderosa ferramenta para os geneticistas, embora a prova definitiva da associação entre mapas de ligação e cromossomos tenha vindo mais tarde, com os estudos realizados por B. McClintock em cromossomos de milho. Creighton; McClintock (1931) demonstraram que o crossing-over é resultado de troca entre segmentos cromossômicos.

Primeiramente, para a construção do mapa genético, são selecionados aqueles marcadores que apresentam polimorfismo. Para isto, realiza-se o teste de segregação na análise individual das marcas por meio da estatística de qui-quadrado (χ^2). O objetivo da triagem dos marcadores genéticos e os testes de segregação é selecionar adequadamente os marcadores que apresentam as frequências mendelianas esperadas de um único loco em uma população. A ocorrência de segregação distorcida significativa

indica que o modelo genético adotado pode ser inapropriado, que os dados são de baixa qualidade ou que o processo de amostragem não foi aleatório (LIU, 1998).

Vale ressaltar que quando se realizam testes conjuntos de segregação para várias marcas, é necessário estabelecer um nível crítico de significância que reduza a probabilidade de rejeição de uma hipótese verdadeira. Assim, recomenda-se a utilização da correção de Bonferroni. Esta correção tem por objetivo fixar um nível de significância α_T para o experimento. Para isso, deve-se estimar o nível de significância α para cada teste que proporcione o nível de significância α_T desejado para o experimento. (SCHUSTER; CRUZ, 2008).

Segundo Alzate-Marin et al. (2005), uma das etapas mais importantes no uso de marcadores moleculares é o estabelecimento da relação entre um dado marcador e um Locus de Características Quantitativas (QTL - Quantitative Trait Loci) de interesse. Esta é uma etapa trabalhosa, sendo um dos motivos desta dificuldade o fenômeno de recombinação entre as regiões que circunvizinham o loco de interesse, que podem ser distintas, mesmo entre materiais genéticos aparentados. Portanto, um marcador polimórfico entre os genitores A e B pode não ser polimórfico entre A e C. Logo, para cada cruzamento, marcadores específicos devem ser identificados. Em muitos casos, no entanto, um mesmo marcador pode ser útil em diferentes populações derivadas de diferentes cruzamentos.

Após a estimação das frequências de recombinação, obtém-se os grupos de ligação. Para isto, primeiramente precisa-se estimar as frequências de recombinação entre os pares de locos. Dois marcadores são considerados ligados quando a frequência de recombinação entre eles for inferior a um limite predefinido e o LOD for superior a um limite também pré-definido. De maneira geral, tem sido comumente adotado como critério valores de r_{\max} igual a 30 centimorgans (cM) e LOD_{\min} igual a 3. As frequências de recombinação podem ser estimadas pelos métodos da verossimilhança (usando algoritmo gráfico, iterativo de Newton-Raphson ou esperança e maximização) e Bayesiano (SCHUSTER; CRUZ, 2008; TOLEDO et al., 2008).

Um importante fato no mapeamento é a falta de aditividade das frequências de recombinação, o que levou ao desenvolvimento das funções de mapeamento. Elas são utilizadas para converter frequências de recombinação, expressas em unidades de mapa (cM), em medidas de distâncias com propriedades mais interessantes para o ordenamento dos locos. As mais conhecidas são as de Haldane (1919) e de Kosambi (1944). A função de Haldane admite a independência das permutas nos intervalos

adjacentes, enquanto a função de Kosambi considera a interferência. Sabe-se que a coincidência é um valor definido no intervalo de $0 \leq c \leq 1$ e que, quanto menor a frequência de recombinação, maior a interferência. Dessa forma, Kossambi (1944) considera que c pode ser expresso por $2r$. Uma vez estimadas as frequências de recombinação entre os pares de locos e os grupos de ligação formados, portanto, verifica-se qual é a melhor ordem das n marcas nos grupos de ligação (SCHUSTER; CRUZ, 2008).

A maioria dos trabalhos encontrados na literatura, referentes a mapeamento, utiliza população de mapeamento do tipo F_2 (HONKATUKIA et al., 2011; MCCLOSKEY et al., 2011; SILVA et al., 2011; YAN et al., 2009; MASOJÉ; MILCZARSKI, 2009; PIRES et al. 2005). Também foram verificados trabalhos com população de RILs (BALASUBRAMANIAN et al., 2009; LISEC et al., 2009; GHANDILYAN et al., 2009), retrocruzamentos (JIANG et al., 2009; EMEBIRI et al., 2009), famílias de irmãos completos (CHIANG et al., 2011; COBAT et al., 2011; ROSADO et al., 2010; PEIXOTO et al. 2009; BHERING; CRUZ, 2008) e famílias de meio-irmãos (PEIXOTO et al. 2009; SILVA et al., 2005).

2.2. Populações de mapeamento

Os QTLs são regiões do genoma que contém poucos ou muitos genes com efeitos genéticos aditivos, dominantes e epistáticos. A identificação dessas regiões permite identificar genes que controlam, parcialmente ou num todo, características fenotípicas (LIU, 1998; PAIXÃO et al., 2008).

Em estudos de mapeamento, visando à detecção de QTL, a estrutura das populações, as informações dos marcadores moleculares e dos fenótipos são de suma importância. Por meio dessas informações, foram desenvolvidas várias metodologias visando o mapeamento e a detecção de QTL. A finalidade dessas metodologias é responder algumas questões como: número de QTL, modo de ação e magnitude do seu efeito, facilitando a manipulação dessas informações para o melhoramento (GONÇALVES et al., 2005).

Para identificação de marcadores ligados às características de interesse é necessária a disponibilidade de populações segregantes. Há várias estruturas de populações oriundas de cruzamentos controlados, sendo que as mais utilizadas são: F_2 ,

retrocruzamento, duplo haplóides e linhagens endogâmicas recombinantes (RILs) (ALZATE-MARIN et al., 2005).

Uma população oriunda de cruzamento controlado pode segregar em diversas proporções, sendo que as mais comuns são (SCHUSTER; CRUZ, 2008):

- a) Segregação 1:1 para os marcadores Mm e mm, com uma probabilidade de obter Mm, ou mm, ao se avaliar uma população é de 0,5, para as populações de retrocruzamento;
- b) Segregação 1:1 para os marcadores MM e mm, com uma probabilidade de obter MM, ou mm, ao se avaliar uma população é de 0,5, para as populações RILs e duplo haplóides;
- c) Segregação 3:1 para os marcadores M₋ e mm, para marcadores dominantes em populações F₂. Neste caso, não é possível diferenciar os genótipos MM e Mm;
- d) Segregação 1:2:1 para os marcadores MM, Mm e mm, para marcadores codominantes em populações F₂. Desta forma diferencia-se os genótipos MM, Mm e mm.

Além das populações de cruzamento controlado, para este tipo de estudo também pode ser utilizado populações exogâmicas. Estas também são conhecidas como populações naturais e podem ser geradas por acasalamentos preferenciais, autofecundação ou por acasalamento ao acaso. Exemplos de populações exogâmicas são as famílias de irmãos completos e de meio-irmãos. Populações geradas de uma única planta polinizada por muitas fontes de pólen desconhecidas ou parcialmente conhecidas são famílias de meio-irmãos e quando se sabe a origem do pólen é denominado família de irmãos completos. Outro tipo de população natural, comum no reino vegetal, é gerado pela combinação de polinização ao acaso e autopolinização (LIU, 1998).

2.2.1. Metodologias para detecção de QTL

2.2.1.1. Método da marca simples

Em populações de cruzamentos controlados há vários procedimentos estatísticos para determinar se um gene está ligado a um loco marcador. A metodologia conhecida como BSA – Bulk Segregant Analysis (ARNHEIM et al., 1985; MICHELMORE et al., 1991), ou análises de bulks segregantes, é normalmente utilizada para detecção de

genes que controlam características de herança mendelianas simples. É considerado o processo mais direto de detecção de um marcador associado a um QTL, pois trata-se de procedimentos baseados na seleção prévia de marcadores, no sentido de otimizar o mapeamento de regiões flanqueando locos de caracteres quantitativos. Este método consiste em estabelecer dois grupos de indivíduos, a partir de pontos extremos de uma população segregante para um caráter quantitativo e, então, avaliar a ocorrência de polimorfismo em locos de marcadores moleculares entre os grupos (TANKSLEY, 1993; SCHUSTER; CRUZ, 2008).

A vantagem da análise de extremos é o menor gasto de tempo e de recursos nas análises de marcadores. As desvantagens desta análise são: necessidade de maior número de indivíduos segregantes a ser analisados para fenótipos quantitativos que possibilitam coletas suficientes nos extremos da distribuição; embora a análise da distribuição de extremos seja eficiente para detecção de ligação entre os locos marcados e o QTL, ela é pouco eficiente na determinação de efeitos individuais dos QTLs; normalmente é impraticável o uso da distribuição dos extremos para mapear mais de um caráter quantitativo quando os indivíduos com fenótipos extremos para um caráter não representam os extremos do outro caráter (TANKSLEY, 1993).

A associação entre marcadores moleculares e QTLs pode ser evidenciada, de forma mais precisa, por meio das seguintes metodologias: Teste t, análise de variância (ANOVA), regressão linear e máxima verossimilhança. A análise de QTLs pelo método teste t é realizada para populações de retrocruzamento, RILs e duplo-haplóide. Por este método, a análise de associação das marcas é realizada por meio de comparação entre médias. Para essas análises não haverá efeito do marcador sobre a característica quando não existir ligação entre o QTL e o marcador, ou seja, $r=0,5$ (SCHUSTER; CRUZ, 2008).

Ghandilyan et. al (2009), com o objetivo de verificar QTLs referentes ao acúmulo de nutrientes em *Arabidopsis thaliana*, fizeram o mapeamento de quatro populações RIL (Landsberg, Eringsboda-1, Kondara e Antwerp). As análises de detecção de QTL foram realizadas por análise de variância (ANOVA) e pelo teste t. Estes mesmos métodos foram utilizados por Prudent et. al (2009).

Para verificar se há existência de ligação entre um marcador e um QTL, por meio da análise de regressão linear, deve-se atribuir escores para o genótipo do marcador e faz-se análise de regressão do valor fenotípico da característica (variável dependente) em relação aos escores (variável independente). O efeito significativo da

regressão poderá indicar existência de associação entre o marcador e a característica, e esta associação é dada pela ligação genética entre o QTL e o marcador. O valor do coeficiente de determinação (R^2) da regressão é a proporção da variação fenotípica da característica, explicada pelo marcador (SCHUSTER; CRUZ, 2008).

Assim, é considerado o modelo:

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

sendo,

Y_j é o valor da característica quantitativa avaliada no j-ésimo indivíduo da população F_2 ;

X_{1j} é o código do marcador para o efeito aditivo (AA = 1, Aa = 0 e aa = -1);

X_{2j} é o código do marcador para o efeito dominante (AA = 0, Aa = 1 e aa = 0);

β_0 é o intercepto (média da característica);

β_1 é o coeficiente de regressão associado ao efeito aditivo;

β_2 é o coeficiente de regressão associado ao efeito dominante; e

ε_j é o erro aleatório manifestado na característica no j-ésimo indivíduo.

Com objetivo de estudar a heterose ao nível metabólico, Lisc et al. (2009) cruzaram duas populações de *Arabidopsis thaliana* (Col-0 e Col 24) e obtiveram uma população de RIL com 369 linhagens. Avaliam-se 369 RILs e 732 retrocruzamentos com os pais 1 e 2 e para a detecção de QTL foi utilizado a metodologia de regressão. Verificam-se vários trabalhos que utilizam este método (MIGNON et al., 2009; SHEN et al., 2009; ETGES et al., 2009; COUTO et al., 2010; RODRIGUES et al., 2010).

Outro método de verificar a ligação entre marcas moleculares e QTL é o da máxima verossimilhança. As propriedades deste método foram estudadas, primeiramente, por Fisher (1921) e é largamente utilizado na estimação de parâmetros em genética quantitativa e na genética humana. No mapeamento genético, este método é empregado tanto para obtenção de estimativas da frequência de recombinação entre dois locos de herança simples quanto para cálculos de estimativas de parâmetros no mapeamento de QTL. Para o mapeamento, por meio do método da máxima verossimilhança, as observações podem ser tomadas de qualquer uma das classes (QQ, Qq e qq) e assume-se que dentro de cada classe a distribuição é normal (CRUZ et al., 2009).

2.2.1.2. Método de estudo de QTL por análise de intervalo

O mapeamento pela metodologia da marca simples, que não utiliza o mapa genético, é eficiente quanto à detecção se há ligação entre a marca molecular e o QTL, entretanto não permite a localização de sua posição. Diante disto, Lander; Botstein (1989) propuseram a metodologia de mapeamento por intervalo (*interval mapping*). A ideia dos autores era utilizar um par de marcadores como unidade de análise, e não um único marcador. Esta é a base da maioria dos métodos de mapeamento de QTLs utilizados atualmente (CRUZ et al., 2009).

Para a realização da análise de mapeamento por intervalo, é necessário conhecer as distâncias entre os marcadores M_1 e M_2 , ou seja, é preciso dispor, antes da análise de associação com o QTL, do mapa de ligação dos marcadores. Se as distâncias no mapa de ligação estiverem em porcentagem de recombinação, elas poderão ser facilmente transformadas, utilizando-se determinada função de mapeamento (CRUZ et al., 2009).

O método desenvolvido por Lander; Botstein (1989) testa a possibilidade de ocorrência de um QTL em uma série de posições dentro de um intervalo. Estes autores utilizaram a metodologia de máxima verossimilhança para estimação dos parâmetros. Embora esta metodologia produza estimativas consistentes dos parâmetros e permita realizar teste de hipóteses, os programas computacionais requeridos para isso necessitam realizar muitas etapas dos métodos numéricos iterativos para maximizar a verossimilhança. Esse método pode se tornar impraticável se o modelo incluir vários QTLs. Em 1992, Halley e Knott, Martinez e Curnow e Moreno-Gonzales apresentaram um modelo de mapeamento baseado em regressão linear (simples e múltipla). A regressão do valor observado do fenótipo a um conjunto de possíveis genótipos possibilita estimar a ocorrência em cada posição do QTL entre dois marcadores. Nesta regressão, cada possível genótipo, para o QTL, possui uma probabilidade de ocorrência em função de sua localização no intervalo (SCHUSTER; CRUZ, 2008).

O modelo de regressão para método de detecção de QTLs por intervalo simples em população F_2 é dado por:

$$Y_j = \mu + ax_j^* + dz_j^* + \varepsilon_j$$

sendo,

Y_j é o valor da característica Y no indivíduo j;

μ é a média da característica da população;
 a é o efeito aditivo do loco que está sendo estudado sobre a característica;
 d é o efeito de dominância do loco que está sendo estudado sobre a característica;
 x_j^* e z_j^* são as variáveis condicionadoras e dependentes dos genótipos dos marcadores que flanqueiam o QTL, no indivíduo j ; e
 ε_j é o erro aleatório.

Outra forma de realizar análise de QTL é por meio das metodologias bayesianas, que são chamadas de mapeamento por intervalo bayesiano - BIM, Bayesian Interval Mapping - (TOLEDO et. al, 2008; ROCHA et al., 2010).

A abordagem bayesiana permite combinar a verossimilhança dos dados fenotípicos com distribuições a priori atribuídas a todas as quantidades desconhecidas no modelo (número, localização no genoma e efeitos genéticos dos QTLs) de forma a fornecer distribuições a posteriori a respeito dessas quantidades. Métodos de mapeamento bayesiano podem incorporar a incerteza relativa ao número desconhecido de QTLs na análise; essa incerteza, no entanto, resulta em complicações na obtenção da amostra da distribuição conjunta a posteriori, uma vez que a dimensão do espaço do modelo pode variar. O método utiliza algoritmo Monte Carlo via Cadeia de Markov com Saltos Reversíveis (MCMC-SR), proposto por Green (1995), é uma excelente ferramenta para explorar a distribuição conjunta a posteriori nesse contexto. A utilização de metodologias bayesianas tem se tornado frequente nas aplicações em Genética, principalmente em mapeamento de QTLs usando marcadores moleculares (SILVA; LEANDRO, 2009).

As análises de detecção de QTL por intervalo simples não considera o restante da informação do genoma, podendo apresentar problemas em sua detecção. Visando corrigir este problema, Zeng (1993) propôs um método que considera outras marcas, além das do intervalo. É importante salientar que a inclusão de marcadores que não estão associados com a característica pode levar a um aumento na variância residual, ao invés de redução. Portanto, o ideal é que se utilize apenas um subconjunto de marcadores selecionados como cofatores (c), que serão utilizados na análise (CRUZ et al., 2009). Este método proposto por Zeng (1993) foi denominado de intervalo composto.

O modelo de regressão para método de detecção de QTLs por intervalo composto em população F_2 é dado por:

$$Y_j = [\mu + ax_j^* + dz_j^*] + \sum_{k \neq i, i+1}^c b_k x_{kj} + \varepsilon_j$$

sendo,

y_j é o valor da característica y no indivíduo j ;

x_{kj} é o escore do marcador k no indivíduo j (AA=1, Aa=0, aa=-1);

b_k é o coeficiente de regressão parcial do fenótipo Y com o k -ésimo marcador;

ε_j é o erro aleatório;

O termo entre colchetes é o mesmo apresentado no modelo de regressão para mapeamento de intervalo simples.

Essas metodologias de detecção de QTLs por intervalo foram amplamente difundidas e hoje são as mais utilizadas nas pesquisas. Verificam-se, na literatura, vários trabalhos que utilizam a metodologia de intervalo simples (BRAS et al., 2011; BOROVSKY; PARAN, 2011; SOUSA et al., 2011; RODRIGUES et al. 2010; CHAKRABORTY; ZENG, 2011) e composto (YUAN et al., 2011; LU et al., 2011; LIU et al., 2011; DING et al., 2011).

2.4. Tamanho de população para detecção de QTL

O tamanho da população é de extrema importância em estudos relacionados a mapeamento e detecção de QTL. A resolução do mapa, a capacidade de se determinar em grupos de ligações e a posição do QTL estão diretamente relacionadas com o número de indivíduos genotipados e de marcadores. Para Young (1994), amostras de tamanho reduzido – menos de 50 indivíduos – proporcionam baixa resolução no mapeamento e tornam difícil a detecção de QTLs de pequeno efeito. No entanto, atualmente verifica-se que a maioria dos trabalhos de detecção de QTL utiliza tamanhos populacionais superiores a este.

Tanto o tamanho de população quanto o número de marcas para representação de cromossomos em grupos de ligação ainda não são bem definidos e há falta de padrão para a análise de dados em trabalhos de mapeamento e detecção de QTL (CRUZ, 2006).

Silva (2005), pesquisando o tamanho ideal para mapeamento em populações RILs por meio de estudo de simulação de dados, avaliou a influência do número de indivíduos na população segregante sobre o mapeamento, o efeito da saturação do genoma por marcas sobre o mapeamento e o número adequado de indivíduos a ser utilizados. Para isso, geraram genomas com níveis de saturação de 5, 10 e 20 cM com 231, 121 e 66 marcadores, respectivamente. Para cada saturação foram geradas populações de tamanhos 50, 100, 154, 200, 300, 500 e 800 indivíduos. Neste trabalho, avaliaram-se o número de grupos de ligação e de marcas por grupo, tamanho de grupo de ligação, distâncias adjacentes, inversão das marcas e grau de concordância das distâncias nos mapas com os originais (dada pelo estresse). Assim, concluiu-se que para obtenção de mapas confiáveis de população RILs deve-se, necessariamente, levar em consideração o tamanho da população e o número de marcas. Resultados confiáveis foram obtidos a partir dos seguintes tamanhos: 100, 154 e 500 indivíduos, nas saturações de 5, 10 e 20 cM, respectivamente.

Trabalho semelhante foi realizado por Bhering; Cruz (2008), que avaliaram o tamanho ótimo de populações em famílias de irmãos completos por meio de simulações de genomas e quatro tamanhos de populações (100, 200, 400 e 600 indivíduos) do tipo completamente informativas e não completamente informativas, com saturação de 10 cM. Para isto, foram avaliados o número e tamanhos de grupos de ligação, distâncias entre marcas adjacentes, variâncias destas distâncias, estresse e inversões nas posições das marcas. Concluíram que, para populações completamente informativas, o tamanho populacional de 200 indivíduos é suficiente para recuperar as informações originais; contudo, para a população não completamente informativa, é necessária a utilização de uma população maior, de 600 indivíduos.

Vale ressaltar que esses estudos demonstram os tamanhos mínimos necessários para a recuperação das informações genômicas relacionadas ao mapeamento. Isso não quer dizer que estes tamanhos são ideais para detecção de QTL nestas populações.

Yan et al. (2009) estudaram, por meio de simulação, diferentes tamanhos de populações F_2 e herdabilidades. Para este trabalho foram simulados três tamanhos de populações (50, 100 e 200 indivíduos) e três diferentes herdabilidades (5%, 10% e 20%). O tamanho dos cromossomos simulado foi de 100 cM, sendo que a posição do

QTL foi a 55 cM. Verificou-se que quanto maior o tamanho da população, maior será o poder de detecção do QTL; o mesmo ocorre com a herdabilidade.

Verificam-se na literatura poucos trabalhos que estudam a eficiência dos tamanhos populacionais para obtenção de mapas de qualidade (SILVA, 2005; BHERING; CRUZ, 2008) e detecção de QTL (YAN et al., 2009). No entanto, são encontrados trabalhos de detecção de QTL utilizando vários tamanhos de população desde inferiores a 100 até mais de 1000 indivíduos, tanto em populações oriundas de cruzamento controlado quanto de populações exogâmicas.

Foram encontrados vários trabalhos visando à detecção de QTLs em populações F_2 com diversos tamanhos de população. Verificaram-se populações com até 200 indivíduos (COUTO et al., 2010; BRAS et al., 2011; SANTOS, 2010; ROCHA et al., 2010; BOROVSKY; PARAN, 2011; LU et al., 2011; YUAN et al., 2011), entre 201 e 500 (RODRIGUES et al., 2010; DING et al., 2011; SILVA et al., 2011) e mais de 500 (TORTEREAU et al., 2011; SOUSA et al., 2011).

Em relação às populações exogâmicas foram verificados diversos tamanhos populacionais, com menos de 100 (TONG et al., 2011; SILVA et al., 2005), entre 101 e 200 (ROSADO et al., 2010), entre 201 e 1000 (ALAM et al., 2011; CHIANG et al., 2011; ZHENG et al., 2006) e mais de 1000 indivíduos (RAMOS et al., 2011; KOING et al., 2010) para estudos de detecção de QTL.

2.5. Detecção de QTLs fantasmas

Em 1989, Lander e Botstein propuseram a metodologia de mapeamento por intervalo. Este método utiliza, para análise, marcadores adjacentes flanqueando um possível QTL, permitindo, dessa forma, o aumento no poder de detecção e precisão nas estimativas dos efeitos e nas posições do QTL. Entretanto, o mapeamento por intervalo, que avalia um intervalo de cada vez sem considerar o restante da informação do genoma, pode apresentar problemas na detecção quando existem outros QTLs de grande efeito no mesmo ou em outros grupos de ligação. Desse modo, não há como determinar se o efeito significativo observado é devido a um ou a vários QTLs. Além disso, quando múltiplos QTLs estão presentes, um problema comum é a observação de vários falsos positivos, ou seja, são detectados QTLs inexistentes, denominados de QTLs fantasmas. A detecção destes QTLs ocorre devido ao erro Tipo I (rejeitar uma hipótese de nulidade verdadeira – não existência de QTL associada à marca); portanto, quanto maior a

magnitude do nível de significância, menor é a credibilidade atribuída à associação entre marcador e QTL (SCHUSTER; CRUZ, 2008; JANGARELLI et al., 2011).

O mapeamento por intervalo simples pode apresentar o problema de identificar QTLs fantasmas. Visando eliminar este efeito, Zeng (1993) propôs métodos que combinam mapeamento por intervalo e regressão linear múltipla, denominado de mapeamento por intervalo composto (CIM). A principal característica destes métodos é evitar que um QTL presente em qualquer posição do genoma interfira no mapeamento do QTL no intervalo considerado (LAPERUTA, 2011).

O número de marcadores utilizados e o tamanho da população avaliada são fatores relevantes para obtenção de associações significativas entre marcadores e QTL. A utilização de poucos marcadores distribuídos aleatoriamente no genoma diminui as chances de encontrar desequilíbrio de ligação entre marcador e QTL, que é fundamental para detectar ligações entre marcadores e locos quantitativos. Já o uso de muitos marcadores e avaliados sob baixo rigor estatístico pode levar à detecção de QTLs fantasmas (JANGARELLI et al., 2010; JANGARELLI; EUCLYDES, 2010).

Silva et al. (2005), utilizando mapeamento por intervalo, em estudos de simulação, não verificaram corretamente a posição dos QTLs nos intervalos considerados. Foram localizados três QTLs fantasmas, sendo um entre os dois simulados, um à esquerda do primeiro e outro à direita do segundo QTL simulado. Quando o número de informações dentro de família aumentou de 10 para 40 progênies, diminuiu o viés na localização dos QTLs. Da et al. (2000) verificaram que a utilização do mapeamento por intervalo pode identificar apenas um QTL em determinado intervalo que contenha dois QTLs. Este problema poderia ser sanado com a utilização da metodologia de intervalo composto.

Verificam-se na literatura trabalhos que utilizam a metodologia de intervalo composto com o intuito de evitar os QTLs fantasmas (XU et al., 2011; HENNING et al., 2011; SU et al., 2010; LI et al., 2010).

3. REFERÊNCIAS BIBLIOGRÁFICAS

ALAM, M.; LEE, Y. M.; PARK, B. L.; KIM, J. H.; LEE, S. S.; SHIN, H. D.; KIM, K. S.; KIM, N. S.; KIM, J. J. A whole genome association study to detect single nucleotide polymorphisms for body conformation traits in a hanwoo population. **Asian-Aust. J. Anim. Sci.**, v.24, n.3, p.322-329, 2011

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, p.333-342, 2005.

ARNHEIM, H.; STRANGE, C.; ELRICH, H. Use of pooled DNA samples to detected linkage disequilibrium of polymorphic restriction fragments and human disease: studies of HLA class II loci. **Proc. Natl. Acad. Sci., USA**, v.82, p.6970-6974, 1985.

BALASUBRAMANIAN, S.; SCHWARTZ, C.; SINGH, A.; WARTHMAN, N.; KIM, M.C.; Maloof, J.N.; LOUDET, O.; TRAINER, G.T.; DABI, T.; BOREVITZ, J.O.; CHORY, J.; WEIGEL, D. QTL mapping in new arabidopsis thaliana advanced intercross-recombinant inbred lines. **Plos One**, v.4, p.1-8, 2009.

BOROVSKY, Y.; PARAN, I. Characterization of *fs 10.1*, a major QTL controlling fruit in elongation in *Capsicum*. **Theor Appl Genet**, 2011. doi: 10.1007/s00122-011-1615-7

BHERING, L.L.; CRUZ, C.D. Tamanho de população ideal para mapeamento genético em famílias de irmãos completos. **Pesquisa agropecuária brasileira**, v.43, p.379-385, 2008.

BRAS, Y. L.; DECHAMP, N.; KRIEG, F.; FILANGI, F.; GUYOMARD, R.; BOUSSAHA, M.; BOVENHUIS, H.; POTTINGUER, T. G.; PRUNET, P.; ROY, P. L.; QUILLET, E. Detect of QTL with effects on osmoregulation capacities in the rainbow trout (*Oncorhynchus mykiss*). **BMC Genetics**, v.12, n.46, 2011. doi: 10.1186/1471-2156-12-46

CHAKRABORTY, S.; ZENG, Z. B. QTL mapping for Days to Flowering under Drought Condition in Rice (*Oryza sativa* L.) Genome. **Not Bot Hort Agrobot Cluj**, v.39, n.1, p.58-63, 2011.

CHIANG, M. C.; BARYSHEVA, M.; TOGA, A. W.; MEDLAND, S. E.; HANSELL, N.K.; JAMES, M. R.; MCMAHAON, K. L.; ZUBICARAY, G. I. de; MARTIN, N. G.; WRIGHT, M. J.; THOMPSON, P. M. BDNF gene effects on brain circuitry replicated in 455 twins. **NeuroImage**, 2011. DOI: 10.1016/j.neuroimage.2010.12.053

COBAT, A.; ABEL, L.; ALCAIS, A. The Maximum-Likelihood-Binomial Method Revised: A Robust Approach for Model Free Linkage Analysis of Quantitative Trait in Large Sibships. **Genetic Epidemiology**, v.35, p.46-56, 2011.

COUTO, K. R.; SANTOS, J. B.; RAMALHO, M. A. P.; SILVA, G. S. Identificação de marcadores microssatélites relacionados ao escurecimento do grão de feijão. **Pesquisa Agropecuária Brasileira**, v.45, n.11, p.1268-1274, 2010.

- CREIGHTON, H.B.; McCLINTOCK, B. A correlation of citological and genetical crossing-over in *Zea mays*. **Proceedings of the National Academic of Sciences of the United States of America**, Washington, v.17, p.492-497, 1931.
- CRUZ, C. D.; BHERING, L. L.; GOD, P. I. V. G. Mapeamento de QTLs em populações derivadas de cruzamentos controlados. In: BORÉM, A.; CAIXETA, E. T. **Marcadores moleculares**. 2ª Ed., ed. UFV, Viçosa, p.483-532, 2009.
- CRUZ, C. D; GOOD GOD, P. I. V.; BHERING, L. L. Mapeamento de QTLs em populações exogâmicas. In: CAIXETA, E. T.; BORÉM, A. **Marcadores moleculares**. 2º Ed., ed. UFV, Viçosa, p.443-481, 2009.
- CRUZ, E. M. **Efeito da saturação e do tamanho de população F₂ e de retrocruzamento sobre a acurácia de mapeamento genético** (2006). Tese (Doutorado em Genética e Melhoramento) – Universidade Federal de Viçosa, Viçosa.
- DA, Y.; Van RADEN, P.M.; SCHOOK, L.B. Detection and parameter estimation for quantitative trait loci using regression models and multiple markers. **Genetics Selection Evolution**, v.32, p.357-381, 2000.
- DA, Y.; LEWIN, H.A. Linkage information content and efficiency of full-sib and half-sib designs for genes mapping. **Theoretical and Applied Genetics**, v.90, p.699-706, 1995.
- DING, J. Q.; MA, J. L.; ZHANG, C. R.; DONG, H. F.; XI, Z. Y.; XIA, Z. L.; WU, J. Y. QTL mapping for test weight by using F_{2:3} population in maize. **Journal of Genetics**, v.90, n.1, p.75-80, 2011.
- EMEBIRI, L.; MICHAEL, E.P.; MOODY, E.D.B.; OGBONNAYA, E.F.C; BLACK, E.C. Pyramiding QTLs to improve malting quality in barley: gains in phenotype and genetic diversity. **Mol Breeding**, v.23, p.219-228, 2009.
- ETGES, W.J.; OLIVEIRA, C.C.; RITCHIE, M.G.; NOO, M.A.F. Genetics of incipient speciation in *Drosophila mojavensis*: ii. host plants and mating status influence cuticular hydrocarbon qtl expression and g × e interactions. **Evolution**, v.63, p.1712-1730, 2009.
- FEITOSA, M.; KRIEGER, H. O futuro da epidemiologia genética de características complexas. **Ciência e Saúde Coletiva**, v.7, p.71-83, 2002.
- FISCHER, R. A. On the mathematical foundations of statistics. **Phil. trans. Royal Soc.**, Londres, v.222, p.309-368, 1921.
- GHANDILYAN, A.; ILK, N.; HANHART, C.; MBENGUE, M.; BARBOZA, L.; SCHAT, H.; KOORNNEEF, M.; EL-LITHY, M.; VREUGDENHIL, D.; REYMOND, M.; AARTS, M.G.M. A strong effect of growth medium and organ type on the identification of QTLs for phytate and mineral concentrations in three *Arabidopsis thaliana* RIL populations. **Research Paper**, v.60, p.1409-1425, 2009.

- GONÇALVES, T.M.; OLIVEIRA, H.N.; BOVENHUIS, H.; BINK, M.; ARENDONK, J.V. Modelos alternativos para detecção de locos de características quantitativas (QTL) de carcaça e crescimento nos cromossomos 4, 5 e 7 de suínos. **Revista Brasileira de Zootecnia**, v.34, p.1540-1552, 2005.
- GREEN, P.J. Reversible jump markov chain monte carlo computation and bayesian model determination. **Biometrics**, v.82, p.711-732, 1995.
- GRIFFITHS, A.J.F.; MILLER, J.H.; SUZUKI, D.T.; LEWONTIN, R.C.; GELBART, W.M. **Introdução à genética**. Rio de Janeiro: Guanabara, 1998. 856p.
- HALDANE, J.B.S. The combination of linkage values, and the calculation of distance between the loci of linked factors. **Journal of Genetics**, v.8, p.299-309, 1919.
- HASEMAN, J. K.; ELSTON, R. C. The investigation of linkage between a quantitative trait and a marker locus. **Behav. Genet.**, v.2, p.3-19, 1972.
- HENNING, J. A.; TOWNSEND, M. S.; GENT, D. H.; BASSIL, N.; MATTHEWS, P.; BUCK, E.; BEATSON, R. QTL mapping of powdery mildew susceptibility in hop (*Humulus lupulus* L.). **Euphytica**, v.180, p.411-420, 2011. DOI: 10.1007/s10681-011-0403-4
- HONKATUKIA, M.; TUISKULA-HAAVISTO, M.; AHOLA, V.; UIMARI, P.; SCHMUTZ, M.; PREISINGER, R.; CAVERO, D.; VANNERSTROM, P.; ARANGO, J.; O'SULLIVAN, N.; FULTON, J.; VILKKI, J. Mapping of QTL affecting incidence of blood and meat inclusion in egg layers. **BMC Genetics**, v.12, n.55, 2011. DOI: 10.1186/1471-2156-12-55
- JANGARELLI, M.; EUCLYDES, R. F.; CECON, P. R. Níveis de significância na identificação de marcadores moleculares no mapeamento genômico. **Revista Brasileira de Zootecnia**, v.40, n.2, p.308-313, 2011.
- JANGARELLI, M.; EUCLYDES, R. F.; CECON, P. R. Desempenho fenotípico ao utilizar diferentes densidades de marcadores moleculares no mapeamento genômico. **Bioscience Journal**, v.26, n.4, p.626-631, 2010.
- JANGARELLI, M.; EUCLYDES, R. F. Estimativas de parâmetros genéticos utilizando diferentes níveis de saturação no mapeamento genômico. **Revista de Ciências Médicas e Biológicas**, v.9, n.1, p.22-28, 2010.
- JANSEN, R. C. Interval mapping of multiple quantitative trait loci. **Genetics**, v.135, p.205-211, 1993.
- JANSEN, R. C. A general Monte Carlo method for mapping multiple quantitative trait loci. **Genetics**, v.192, p.305-311, 1996.
- JIANG, F.; YOSHIDA, T.; NAKAKI, F.; TERAWAKI, S.; CHIKUMA, S.; OKAZAKI, I.; HONJO, T.; OKAZAKI, T. Identification of QTLs that modify peripheral neuropathy in NOD.H2^b-Pdc1^{-/-} mice. **International Immunology**, v.21, p.499-509, 2009.

- KOSAMBI, D.D. The estimation of map distance from recombination values. **Annuaire of Eugenetics**, v.12, p.172-175, 1944.
- LANDERS, S.; BOTSTEIN, D. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. **Genetics**, v.121, n.1, p.185-199, 1989.
- LANDER, E.S.; WEINBERG, R.A. Genomics: journey to the century of biology. **Science**, v.287, p.1777-1782, 2000.
- LAPERUTA, L. D. C. Estudo de uma população segregante (F₁) de maracujá doce: enriquecimento do mapa de ligação e mapeamento de QTL para produção e qualidade de frutos. **Tese de doutorado** (Apresentada à Universidade de São Paulo/ESALQ no programa de pós-graduação em Ciências). 134 f. 2011.
- LI, H.; HEARNE, S.; BÄNZIGER, M.; LI, Z.; WANG, J. Statistical properties of QTL linkage mapping in bilateral genetic population. **Heredity**, v.105, p.257-267, 2010. DOI: 10.1038/hdy.2010.56
- LISEC, J.; STEINFATH, M.; MEYER, R.C.; SELBIG, J.; MELCHINGER, A.E. WILLMITZER, L.; ALTMANN, T. Identification of heterotic metabolite QTL in *Arabidopsis thaliana* RIL and IL populations. **The Plant Journal**, v.59, p.777-788, 2009.
- LIU, B.H. **Statistical genomic: linkage, mapping, and QTL analysis**. Boca Raton: CRC Press. 1998. 611p.
- LIU, R.; WANG, B.; GUO, W.; WANG, L.; ZHANG, T. Differential gene expression and associate QTL mapping for cotton yield based on a cDNA-AFLP transcriptome map in an immortalized F₂. **Theor Appl Genet**, DOI 10.1007/s00122-011-1597-5, 2011.
- LU, B.; XIE, K.; YANG, C.; ZHANG, L.; WU, T.; LI, L.; LIU, X.; JIANG, L.; WAN, J. Efficient QTL detection for heading date in bechcross inbred line and F₂ population derived from the same rice cross. **African Journal of Agriculture Research**, v.6, n.10, p.2372-2378, 2011.
- LYNCH, M.; WALSH, B. **Genetics and Analysis of Quantitative Traits**. Sunderland, MA: Sinauer Associates, 1998. 980 p.
- KOING, I. R.; SCHUMACHER, J.; HOFFMANN, P.; KLEENSANG, A.; LUDWIG, K. U.; GRIMM, T.; NEUHOFF, N.; PREIS, M.; ROESKED.; WARNKE, A.; PROPPING, P.; REMSCHMIDIT, H.; NOTHEN, M. M.; ZIEGLE, A.; MULLER-MYHSOK, B.; SCHULTE-KORNE, G. Mapping for dyslexia and related cognitive trait loci provides strong evidence for further risk genes on chromosome 6p21. **American Journal of Medical Genetics**, v.156, p.36-40, 2010. doi: 10.1002/ajmg.b.31135
- MARTINEZ, M.L.; VUKASINOVIC, N. Algoritmo para cálculo da proporção de genes idênticos por descendência, para mapear QTL em famílias de meio-irmãos. **Revista Brasileira de Zootecnia**, v.29, p.443-451, 2000.

MCCLOSKEY, B.; MA, X.; TANKSLEY, S.D. Quantifying the Relative Contribution of the Heterozygous Class to QTL Detect Power. **Statistical Application in Genetics and Molecular Biology**, v.10, n.1, 2011. doi: 10.2202/1544-6116.1622

MICHELMORE, R.W.; PARAN, L.; KASSELLI, R.V. Identification of markers linked to disease-resistance genes by bulked segregate analysis: a rapid method to detect markers in specific genomic regions by using segregating population. **Proc. Natl. Acad. Sci.**, v.88, p.9828-9832, 1991.

MIGNON, G.L.; PITEL, F.; GILBERT, H.; BIHAN-DURVAL, E.L.; VIGNOLES, F.; DEMEURE, O.; LAGARRIGUE, S.; SIMON, J.; COGBURN, L.A.; AGGREY, S.E.; DOUAIRE, M.; ROY, P.L. A comprehensive analysis of QTL for abdominal fat and breast muscle weights on chicken chromosome 5 using a multivariate approach. **Animal Genetics**, v.40, p.157-164, 2009.

MOSAJÉ, P.; MILEZARSKI, P. Relationship between QTLs for preharvest sprouting and alpha-amylase activity in rye grain. **Mol Breeding**, v.23, p.75-84, 2009.

PAIXÃO, D.M.; GUIMARÃES, S.E.F.; SILVA FILHO, M.I.; LOPES, P.S.; PEREIRA, M.S.; SOLERO, B.P. Detecção de locos de características quantitativas nos cromossomos 16, 17 e 18 de suínos. **Revista Brasileira de Zootecnia**, v.37, p.1781-1787, 2008.

PEIXOTO, M.G.C.D.; MARTINEZ, M.L.; TOLEDO, R.L.; MACHADO, M.A.; CARVALHOR, M.R.S. Detecção de QTL em dados de famílias estruturadas como as de um núcleo MOET por meios de métodos de regressão. **Arq. Bras. Med. Vet. Zootec.**, v.61, p.941-948, 2009.

PIRES, A.V.; LOPES, P.S.; GUIMARÃES, S.E.F.; GUIMARÃES, C.T.; GOMIDE, L.A.M.; BENEVENUTO JÚNIOR, A.A.; CARMO, F.M.S. Quantitative trait loci mapping for meat quality traits in swine chromosome 6. **Arq. Bras. Med. Vet. Zootec.**, v.57, p.608-615, 2005.

PRUDENT, M.; CAUSSE, M.; GÉNARD, M.; TRIPODI, P.; GRANDILLO, S.; BERTIN, N. Genetic and physiological analysis of tomato fruit weight and composition: influence of carbon availability on QTL detection. **Journal of Experimental Botany**, v.60, p.923-937, 2009.

RAMOS, A. M.; DUIJESTEIJN, N.; KNOL, E. F.; MERKS, J. W.; BOVENHUIS, H.; CROOIJAMANS, R. P.; GROENEM, M. A.; HARLIZIUS, B. The distal and of porcine chromosome 6p is involved in the regulation of skatole levels in boars. **BMC Genetics**, v.12, 2011. doi: 10.1186/1471-2156-12-35

ROCHA, R. B.; BARROS, W. S.; MURO-ABAD, J.I.; TOMAZ, R. S.; CRUZ, C. D.; BARROS, E. G.; ARAÚJO, E. F. Método para mapeamento de loco controladores de características oligogênicas. **Ciência Rural**, v.40, n.2, p.332-338, 2010.

RODRIGUES, J. I. S.; MIRANDA, F. D.; FERREIRA, A.; BORGES, L. L.; FERREIRA, M. F. S.; GOOD-GOD, P. I. V.; PIOVESAN, N. D.; BARROS, E. G.; CRUZ, C. D.; MOREIRA, M. A. Mapeamento de QTL para conteúdo de proteína e óleo em soja. **Pesquisa Agropecuária Brasileira**, v.45, n.5, p.472-480, 2010.

ROSADO, T. B.; TOMAZ, R. S.; RIBEIRO JUNIOR, M. F.; ROSADO, A. M.; GUIMARÃES, L. M. S.; ARAUJO, E. F.; ALFENAS, A. C.; CRUZ, C. D. Detection of QTL associated with roset resistance using IDB-based methodologies in exogamic *Eucalyptus* ssp. population. **Crop Breeding and Applied Biotechnology**, v.10, p.321-328, 2010.

SANTOS, R. F. **Desenvolvimento de ferramentas genéticas e genômicas para introgressão de genes silvestres no amendoim cultivado**. Brasília: Faculdade de Agronomia e Medicina Veterinária, Universidade de Brasília, 2010, 44p. Dissertação de Mestrado.

SCHUSTER, I.; CRUZ, C.D. **Estatística genômica aplicada a populações derivadas de cruzamento controlados**. 2° Ed. Viçosa, 2008. 568p.

SHEN, Q.; WANG, X.; CHEN, Y.; XU, L.; WANG, X.; LU, L. Expression QTL and regulatory network analysis of microtubule-associated protein tau gene. **Parkinsonism and Related Disorders**, v.15, p.525-531, 2009.

SIEGMUND, D.; YAKIR, B. **The statistics of gen mapping**. New York, NY: Springer, 2007. 331 p.

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

SILVA, H. D.; VENCOVSKY, R. Poder de detecção de “Quantitative Trait Loci”, da análise de marca simples e da regressão linear múltipla. **Scientia Agricola**, v.59, n.4, p.755-762, 2002.

SILVA, F. E. J.; GUIDO, L. N.; GAZAFFI, R.; GARCIA, A. A. F.; LEDUR, M. C.; COUTINHO, L. L.; ROSÁRIO, M. F. Regiões genômicas associadas a características de desempenho e carcaça no cromossomo 5 de linhagens brasileiras de galinhas. **Pesquisa Agropecuária Brasileira**, v.46, n.3, p.229-238, 2011.

SILVA, J.P.; LEANDRO, R.A. Uma abordagem bayesiana para o mapeamento de QTLs utilizando método MCMC com saltos reversíveis. **Ciência e Agrotecnologia**, v.3, p.1061-1070, 2009.

SILVA, M. V. G. B.; MARTINEZ, M. L.; TORRES, R. A.; LOPES, P. S.; EUCLIDES, R. F.; PEREIRA, C. S.; MACHADO, M. A.; ARBEX, W. Modelos aleatórios na estimação da localização de QTLs em famílias de meio-irmãos. **Revista Brasileira de Zootecnia**, v.34, n.1, p.66-75, 2005.

SOUSA, K. R. S.; GUIMARÃES, S. E. F.; SILVA FILHO, M. I.; LOPES, M. S.; PINTO, A. P. G.; VERANDO, L. L.; BRACCINI NETO, J.; LOPES, P. S. Mapeamento de locos de características quantitativas nos cromossomos 5, 7 e 8 de suínos. **Revista Brasileira de Zootecnia**, v.40, n.1, p.115-123, 2011.

STALE, J.; Quantitative trait loci mapping in natural populations: progress, caveats and future directions. **Molecular Ecology**, v.14, p.363-379, 2005.

- SU, C. F.; LU, W. G.; ZHAO, T. J.; GAI, J. Y. Verification and fine-mapping of QTLs conferring days to flowering in soybean using residual heterozygous lines. **Chinese Science Bulletin**, v.55, n.6, p.499-508, 2010. DOI: 10.1007/s11434-010-0032-7
- TANKSLEY, S.D. Mapping polygenes. **Annual Reviews of Genetic**, v.27, p.205-233, 1993.
- TOLEDO, E.R.; LEANDRO, R.A.; SOUZA JUNIOR, C.L.; SOUZA, A.P. Mapeamento de QTLs : uma abordagem bayesiana. **Rev. Bras. Biom.**, v.26, p.107-114, 2008.
- TONG, C.; WANG, Z.; ZHANG, B.; SHI, J.; WU, R. 3FunMap: Full-sib family functional mapping of dynamics traits. **Bioinformatics**, v.27, 14, p.2006-2008, 2011. doi: 10.1093/bioinformatics/btr306
- TORTEREAU, F.; GILBERT, H.; HEUVEN, H. C. M.; BIDANEL, J. P.; GROENEM, M. A. M.; RIQUET, J. Number and mode of inheritance of QTL influencing backfat thickness on SSC2p in sino-european pig pedigrees. **Genetics Section Evolution**, v.43, n.11, 2011. doi: 10.1186/1297-686-43-11
- XU, M. H.; WEI, C. S.; TANG, X. T.; ZHU, Z. H.; SIMA, Y. F.; LOU, X. Y. A new mapping method for quantitative trait loci of silkworm. **Genetics**, 2011. DOI: 10.1186/1471-2156-12-19
- ZENG, Z. B. Theoretical basis of precision mapping of quantitative trait loci. **Proceedings of the National Academy of Science United States of America**, v.90, p.10972-10976, 1993.
- ZENG, Z. B. Precision mapping of quantitative trait loci. **Genetics**, v.136, p.1457-1468, 1994.
- ZENG, G.; GHOSH, K.; CHEN, Z.; LI, Z. Extreme rank section for linkage analysis of quantitative trait loci using selected sib-pairs. **Annals of Human Genetics**, v.69, p.1-10, 2006. doi: 10.1111/j.1529-8817.2006.00268.x
- YAN, A.; ZHI-QIU, H.; ZAI-XIANG, T.; XUE-FENG, W.; CHEN-WU, X. A General Method for QTL Mapping in Multiple Related populations Derived from Multiple Parents. **Rice Science**, v. 16, n. 1, 2009.
- YOUNG, N. D. **Constructing a plant genetic linkage map with DNA markers**. In: PHILIPS, P. L.; VASIL, I. K. DNA-Based Markers in plants. Dordrecht, The Netherlands: Kluwer Academic Publisher, p.39-57, 1994.
- YUAN, J.; WANG, B.; WU, L. Quantitative trait loci (QTL) mapping for inflorescence length traits in *Lablab purpureus* (L.) sweet. **African Journal of Biotechnology**, v.10, n.18, p.3558-3566, 2011.

CAPÍTULO 1

TAMANHO DE POPULAÇÃO IDEAL PARA DETECÇÃO DE QTL EM POPULAÇÕES F_2

RESUMO

FERREIRA, Eliel Alves, D.Sc., Universidade Federal de Viçosa, fevereiro de 2012. **Tamanho de população ideal para detecção de QTL em populações F₂**. Orientador: Cosme Damião Cruz. Coorientadores: Pedro Crescêncio Souza Carneiro e Leonardo Lopes Bhering.

O tamanho da população é de extrema importância em estudos relacionados a mapeamento e detecção de QTL, uma vez que a resolução do mapa, a capacidade de se determinar em grupos de ligações e a detecção de QTL estão diretamente relacionadas com o número de marcadores utilizados e o de indivíduos genotipados. Do exposto, o presente trabalho teve por objetivo o estudo de populações F₂, em relação ao tamanho populacional, visando à detecção de QTLs. Os dados foram gerados utilizando o módulo de simulação do aplicativo computacional GQMOL. Foram simulados genomas parentais e amostras de populações F₂. As amostras formadas foram de quatro tamanhos diferentes: 200, 400, 600 e 1000 indivíduos. Para cada tamanho populacional foram simuladas 50 repetições com as mesmas características. O genoma dessas populações foi simulado tomando como referência uma espécie fictícia diplóide com $2n = 2x = 10$ cromossomos, composto de cinco grupos de ligações. Cada grupo de ligação formado possuiu tamanho de 100 cM, contendo 21 marcadores codominantes com saturação de 5 cM. Estipulou-se que os grupos de ligação 1, 2, 3 e 4 conteriam um QTL, explicando 40%, 30%, 20% e 10% da variação genotípica, respectivamente e que o grupo de ligação 5 não conteria QTL, sendo utilizado como controle da metodologia aplicada à detecção. Para cada repetição foi simulada uma característica com média igual a 100, herdabilidade (h^2) igual a 50% e coeficiente variação ambiental (C_{Ve}) igual a 10%. O grau médio de dominância foi considerado igual a 1, ou seja, dominância completa. Posteriormente, foram realizados os testes de segregação e as análises de detecção de QTL. Para este foram utilizadas as metodologias de marca simples, intervalo simples e intervalo composto, quando necessário. Observou-se que quanto menor for o efeito do QTL maior deverá ser a população de mapeamento. No grupo de ligação 1 praticamente não houve diferenças entre os tamanhos populacionais; já no grupo 4 houve diferenças no posicionamento dos QTLs, nos diferentes tamanhos de população. Assim, conclui-se que as populações de 200, 400, 600 e 1000 indivíduos são eficientes para detectar e posicionar QTL que explica acima de 40%, 30%, 20% e 10% da variação genotípica da característica, respectivamente.

Palavras chave: Característica quantitativa; tamanhos populacionais; dados simulados.

ABSTRACT

FERREIRA, Eliel Alves, D.Sc., Universidade Federal de Viçosa, february de 2012. **The best population size for QTLs detection in F₂ population.** Adviser: Cosme Damiano Cruz. Co-advisers: Pedro Crescencio Souza Carneiro and Leonardo Lopes Bhering.

The population size is very importance in studies of mapping and QTL detection, since the map resolution, the ability to determine in linkage groups and QTL detection are related to mark number used and the individuals genotyped. The objective this search was studies the best population size for QTLs detection in F₂ populations. The genomes parental and F₂ populations were simulated using the simulation module of GQMOL software. Were simulate four different population sizes: 200, 400, 600 and 1000 individuals and 50 replicas. The genome was simulated for diploid species with $2n = 2x = 10$ chromosomes, composed of five linkage groups. Each linkage group was of 100 cM, containing 21 co-dominant markers with saturation of 5 cM. The linkage groups 1, 2, 3 and 4 contain a QTL, explaining 40%, 30%, 20% and 10% of phenotypic variation, respectively, and the linkage group 5 does not contain QTL. Each replica was simulated with yield of 100, heritability (h^2) of 50% and coefficient environmental variation (CVe) of 10%. The average degree of dominance was considered equal to 1. Later, were applied segregation tests and QTL analysis. For QTL analysis was used the simple marking methodologies, simple and composite interval. It was observed that the smaller effect the higher should be population mapping. In linkage group 1, there was no difference, so can use the smaller population. However, in the linkage group 4, QTL explained 10%, there were large differences in the positioning of the QTLs and the population of 1000 individuals showed the less variation and closer to the simulated position, which was 50 cM. It is concluded that the population of 1000 individuals was more efficient in detecting and positioning QTL.

Key words: quantitative trait, population sizes, simulated data.

1. INTRODUÇÃO

A eficiência do melhoramento, tanto vegetal quanto animal, é aumentada com a utilização de mapeamento genético, uma vez que marcas no genoma podem ser associadas a um ou mais genes controladores de uma característica, podendo ser qualitativa ou quantitativa. Sabendo a localização do gene ou QTL de interesse, pode-se transferir esta região para os descendentes. Dessa forma, com o genoma da espécie mapeada, o trabalho de melhoramento poderá ser otimizado, tornando o programa de melhoramento mais eficiente.

A análise Genômica, principalmente as análises de detecção de QTL, é de extrema importância no melhoramento. Com esta análise consegue-se identificar, posicionar e mapear o QTL de interesse e, dessa forma, possibilita a realização do melhoramento assistido por marcadores moleculares. Para que o melhoramento assistido tenha sucesso é importante que a detecção e o posicionamento deste QTL sejam confiáveis, ao contrário pode-se perder tempo e recursos financeiros. A eficiência na identificação e no posicionamento do QTL depende, diretamente, da saturação do genoma, do tamanho da população e da metodologia aplicada à detecção de QTL.

Verificam-se trabalhos relatando que amostras de tamanho reduzido, inferiores a 50 indivíduos, proporcionam baixa resolução no mapeamento e tornam difícil a detecção de QTLs de pequeno efeito. No entanto, atualmente, o tamanho das populações utilizadas são maiores, sendo encontrados trabalhos que estudam a influência do tamanho populacional na formação dos grupos de ligação tanto para cruzamento controlado (F_2 , RILs, retrocruzamento, etc.) quanto para populações exogâmicas (populações de meio-irmãos e de irmão completos).

O tamanho das populações e o número de marcas para representação em grupos de ligação não têm sido bem definidos e, de certa forma, existe falta de padrão para análise de dados nos trabalhos de mapeamento (CRUZ, 2006). Verifica-se que em populações RILs um tamanho de 100 indivíduos com uma saturação de 5 cM recupera os grupos de ligação. Estes grupos também são recuperados em populações de irmãos completos, completamente informativo com um tamanho de 200 indivíduos (SILVA, 2005; BHERING; CRUZ, 2008). É importante ressaltar que tais trabalhos, por meio de simulação, verificam que estes tamanhos populacionais recuperam os grupos de ligação, no entanto, não se sabe se estes tamanhos detectam e posicionam os QTLs.

Do exposto, o presente trabalho teve por objetivo o estudo de populações F_2 visando determinar o melhor tamanho populacional para a detecção, posicionamento, contribuição e efeitos do QTL.

2. MATERIAL E MÉTODOS

2.1. Simulação

2.1.1. Simulação dos dados

Para gerar os dados, foi utilizado o módulo de simulação do aplicativo computacional GQMOL (CRUZ, 2008), que permite estabelecer parâmetros e gerar informações sobre os genomas, genótipos, genitores, indivíduos de diferentes tipos de populações e dados de características quantitativas. Foram simulados genomas parentais e amostras de populações de cruzamento controlado (F_2), sendo que as amostras formadas foram de quatro tamanhos diferentes: 200, 400, 600 e 1000 indivíduos.

Para cada tamanho populacional foram simuladas 50 repetições, com as mesmas características. O genoma dessas populações foi simulado tomando como referência uma espécie fictícia diplóide com $2n = 2x = 10$ cromossomos, composto de cinco grupos de ligação (GL). Cada grupo de ligação formado possuiu tamanho de 100 cM, contendo 21 marcadores codominantes com distância de 5 cM, equidistantes, em cada intervalo.

2.1.2. Simulação dos genitores

Para a simulação dos genitores foram considerados pais homozigotos contrastantes e os mapas genéticos utilizados para obtenção das populações F_2 estão representados na Figura 1.

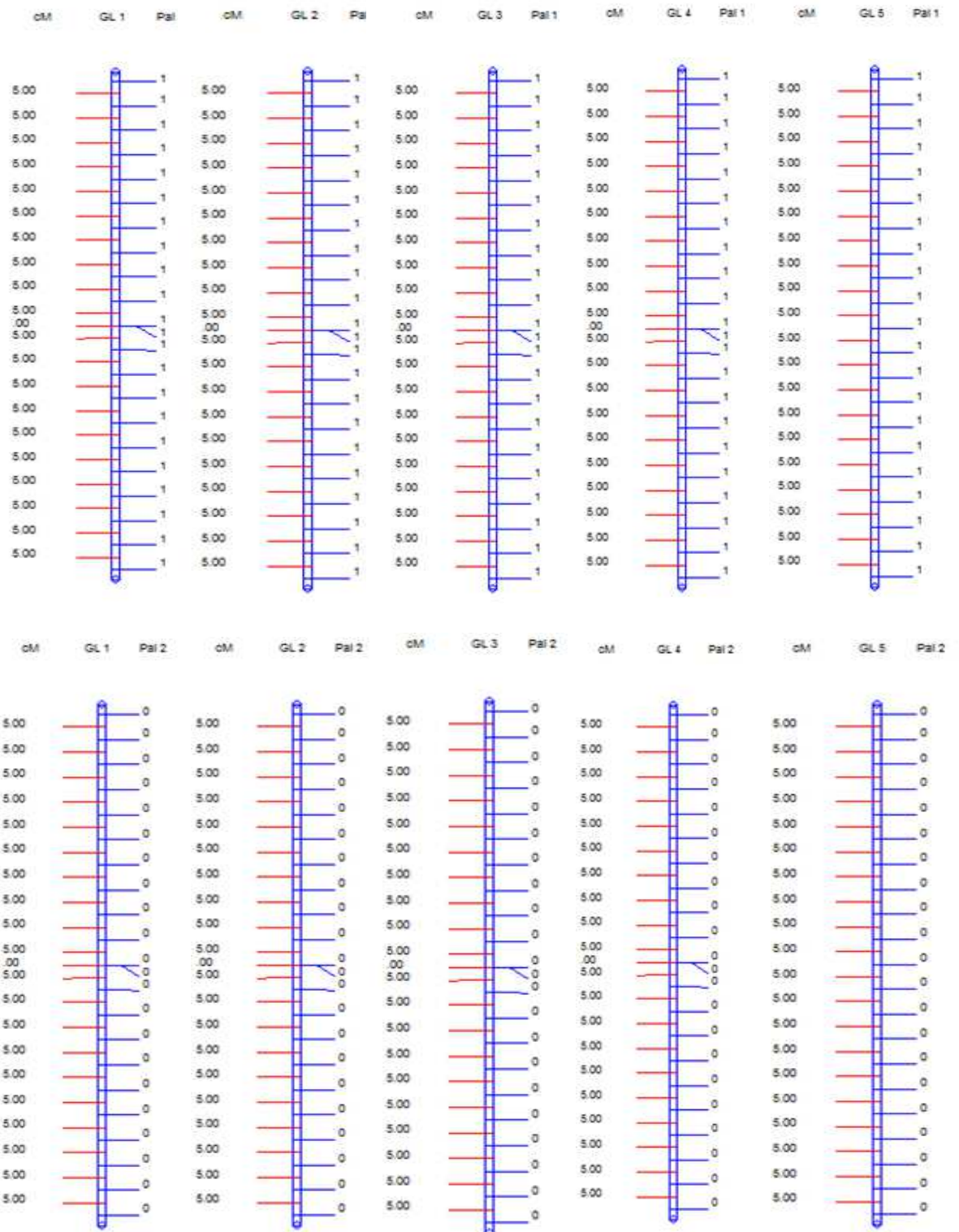


Figura 1. Mapas genéticos com representação genotípica dos genitores utilizados para gerar as populações F_2 a serem analisadas (1=AA e 0=aa).

2.1.3. Procedimentos para simulação dos indivíduos F_2

A estratégia básica de simulação é caminhar ao longo dos cromossomos, realizando permutas em cada intervalo entre marcas adjacentes, de acordo com as distâncias dos marcadores, conforme descrito por Silva (2005).

O processo de simulação das populações F_2 seguiu os seguintes passos:

1. A partir do genoma simulado foram construídos os genótipos parentais, genitores homozigotos contrastantes sempre de padrão genotípico $A_iA_i \times a_ia_i$.
2. A partir dos genótipos parentais foram gerados os gametas, para a formação dos indivíduos da população F_2 . Para formação dos gametas, simulou-se o pareamento dos homólogos, realizando-se permutas ao longo do cromossomo. Considerou-se interferência nula nas regiões delimitadas por dois marcadores adjacentes. A probabilidade de ocorrência de recombinação numa região entre marcadores adjacentes é dada de acordo com a distância destes marcadores no genoma simulado. Sendo que uma maior distância implica numa maior possibilidade de ocorrência de recombinação.

O programa GQMOL considera o encontro aleatório de gametas para a simulação dos indivíduos. Sendo assim, um novo processo acontece para cada indivíduo simulado dentro de cada repetição.

2.1.4. Simulação da característica quantitativa (QTL)

Foi estipulado que os grupos de ligação 1, 2, 3 e 4 contivessem um QTL. Também foi estabelecido que o grupo de ligação 5 não contém QTL, sendo utilizado como controle da metodologia aplicada à detecção de QTL. Para cada repetição foi simulada a característica com média igual a 100, herdabilidade (h^2) igual a 50% e coeficiente variação ambiental (CVe) igual a 10%.

Considerou-se também que a contribuição genética dos QTLs foram diferentes em cada GL, sendo os QTLs posicionados nos GL 1, 2, 3 e 4 explicariam 40%, 30%, 20% e 10% da variação genotípica, respectivamente. O grau médio de dominância foi considerado igual a 1, ou seja, dominância completa.

Os mapas genéticos, com o posicionamento dos QTLs simulados, estão apresentados na Figura 2.

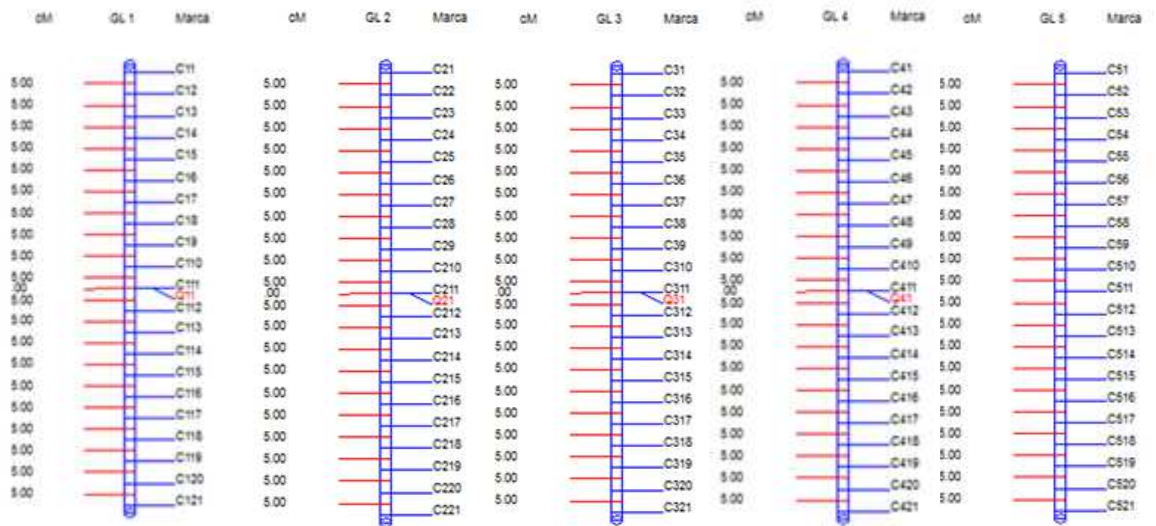


Figura 2. Mapa genético dos cinco grupos de ligação da população F_2 simulada com o posicionamento dos QTLs estabelecidos.

2.2. Mapeamento

Após as populações serem geradas, seguiram-se as etapas do processo de mapeamento, como descrito a seguir.

2.2.1. Análise de segregação

Foram aplicados testes de qui-quadrado (χ^2) para verificar se há segregação em cada marca de todas as populações, sendo esta de 1:2:1.

A estatística utilizada para o teste de segregação foi a qui-quadrado, dada por:

$$\chi^2 = \sum_{i=1}^n \left[\frac{(Obs_i - Esp_i)^2}{Esp_i} \right]$$

sendo,

χ^2 é o valor de qui-quadrado calculado;

Obs_i e Esp_i são valores observados e esperados, para i -ésima classe fenotípica ($i = 1, 2$ e 3).

A hipótese (H_0) de segregação dos marcadores foi testada ao nível conjunto de 5% de probabilidade e dois graus de liberdade. Nas situações em que o valor da probabilidade calculada foi inferior ao pré-estabelecido, a hipótese H_0 foi rejeitada, significando que a segregação não ocorreu de acordo com o esperado.

Como foi testado um grande número de marcadores, foi utilizado a proteção de Bonferroni para reduzir a probabilidade de se cometer o erro tipo I, mantendo a significância total α_T , igual a 5%. Com o uso da correção de Bonferroni o nível de significância para teste individual é dada por:

$$\alpha = -\exp\left(\frac{\ln(1 - \alpha_T)}{N}\right) + 1$$

A utilização da proteção de Bonferroni tem por objetivo fixar um nível crítico de significância α_T para o experimento. Para isso, deve-se estimar um nível de significância α para cada teste que proporcione o nível de significância α_T desejado para cada experimento (CRUZ; SILVA, 2009).

2.2.2. Estimação da porcentagem de recombinação

Após a aplicação dos testes de segregação, seguiu-se a etapa da estimação da porcentagem de recombinação entre os pares de marcas, utilizando-se o método da máxima verossimilhança. Os estimadores de verossimilhança foram obtidos pelo método gráfico em que valores da razão de verossimilhança (LR) ou LOD são analisados graficamente, representativos do eixo Y, associados aos valores da porcentagem de recombinação, no eixo X. A função de verossimilhança é dada por:

$$L(r; n_{ij}) = \lambda p_1^{n_{11}} p_2^{n_{12}} p_3^{n_{13}} p_4^{n_{21}} p_5^{n_{22}} p_6^{n_{23}} p_7^{n_{31}} p_8^{n_{32}} p_9^{n_{33}}$$

sendo,

r é a porcentagem de recombinação;

$$\lambda = \frac{N!}{n_{11}! n_{12}! \dots n_{33}!};$$

$$N = \sum_{i=1}^3 \sum_{j=1}^3 n_{ij};$$

P_i é a probabilidade associada a cada classe;

n_{ij} é o número de indivíduos observados em cada probabilidade.

As probabilidades associadas a cada classe, cujas fases de ligação podem ser de aproximação ou repulsão, juntamente com o número de indivíduos observados estão apresentados na Tabela 1.

Tabela 1. Frequências genotípicas esperadas na população F₂, utilizando marcadores codominantes no cruzamento entre indivíduos com genitores homocigotos contrastantes.

Genótipo	Fase de aproximação			Fase de repulsão		
	Código*	Núm. obs.	Frequência esperada	Código*	Núm. obs.	Frequência esperada
AABB	0-0	n ₁₁	p ₁ =1/4(1-r) ²	0-2	n ₁₃	p ₁ =1/4r ²
AABb	0-1	n ₁₂	p ₂ =1/2r(1-r)	0-1	n ₁₂	p ₂ =1/2r(1-r)
AAbb	0-2	n ₁₃	p ₃ =1/4r ²	0-0	n ₁₁	p ₃ =1/4(1-r) ²
AaBB	1-0	n ₂₁	p ₄ =1/2r(1-r)	1-2	n ₂₃	p ₄ =1/2r(1-r)
AaBb	1-1	n ₂₂	p ₅ =1/2(1-r)+1/2r ²	1-1	n ₂₂	p ₅ =1/2(1-r)+1/2r ²
Aabb	1-2	n ₂₃	p ₆ =1/2r(1-r)	1-0	n ₂₁	p ₆ =1/2r(1-r)
aaBB	2-0	n ₃₁	p ₇ =1/4r ²	2-2	n ₃₃	p ₇ =1/4(1-r) ²
aaBb	2-1	n ₃₂	p ₈ =1/2r(1-r)	2-1	n ₃₂	p ₈ =1/2r(1-r)
aabb	2-2	n ₃₃	p ₉ =1/4(1-r) ²	2-0	n ₃₁	p ₉ =1/4r ²

*Codificação numérica de cada indivíduo, para os dois locos.

O LOD score é um teste de razão de verossimilhança que utiliza o logaritmo na base 10 (Log₁₀) da razão entre a verossimilhança, considerando o valor de r estimado e a verossimilhança para r igual a 0,5.

$$LOD = \text{Log}_{10} \left(\frac{L(\hat{r}; n_{ij})}{L(r=0,5; n_{ij})} \right)$$

Adotou-se, no presente trabalho, o LOD mínimo de 3, indicando que a verossimilhança, considerando o valor estimado de r, deve ser 1000 vezes maior que a verossimilhança em caso de ligação.

2.2.3. Determinação dos grupos de ligação

Na formação de grupo de ligação utilizou-se a propriedade transitiva, ou seja, se o loco A está ligado ao loco B, e o loco B está ligado ao loco C, logo o loco A está ligado ao loco C, independentemente da frequência de recombinação A e C, portanto, A, B e C pertencem ao mesmo grupo de ligação. Os critérios utilizados no agrupamento foram a frequência máxima de recombinação (r_{\max}) e o LOD mínimo (LOD_{\min}), para inferir se dois locos estão ligados. Foram utilizados os valores 30% e 3, respectivamente, para r_{\max} e LOD_{\min} . As marcas mais próximas, em relação às marcas já consideradas, que atenderem aos dois critérios adotados, serão incorporadas ao grupo de ligação. Assim, o processo continua, investigando quanto à porcentagem de recombinação e ao LOD entre marcas e possíveis vizinhos a serem incorporados às extremidades do grupo de ligação em formação.

2.2.4. Ordenamento das marcas no grupo de ligação

Após a formação dos grupos de ligação foi detalhada a melhor ordem, das n marcas nos grupos.

Um dos processos é gerar todas as possíveis ordens ($n!/2$) e adotar como critério a identificação da melhor ordem, sendo aquela que proporcionar menor soma entre distâncias. Contudo, na medida em que o número de locos marcadores aumenta, aumenta-se grandemente o número de ordens possíveis, consumindo, portanto, um grande tempo de processamento. Então, foi utilizado, alternativamente, o critério da soma das frações de recombinação adjacente (SARF – Sum of Adjacent Recombination Fraction) para os mapas construídos sem análise multiloco. Neste processo, a melhor ordem é aquela que apresenta menor soma de recombinações adjacentes. Considera-se a ordem original estabelecida pelo processo de agrupamento e aplica-se o algoritmo RCD (Rapid Chain Delineation), que consiste em realizar permutas entre dois marcadores vizinhos ou distantes envolvendo três ou quatro marcadores. A ordem é alterada se, após a permuta, a soma das distâncias adjacentes for reduzida. Após todas as permutas conclui-se que a melhor ordem é aquela de menor soma de distâncias adjacentes (Schuster & Cruz, 2008).

2.3. Análise de detecção de QTLs

Os métodos utilizados neste trabalho, para detecção de QTL, foram: os métodos da marca simples, intervalo simples e intervalo composto. Todas essas análises foram realizadas utilizando o aplicativo computacional GQMOL (CRUZ, 2008) que pode ser encontrado no endereço <http://www.ufv.br/dbg/biodata.html>.

2.3.1. Método da Marca simples

O método da marca simples foi empregado para avaliar a associação entre genótipos e marcadores moleculares, utilizando-se regressão linear. Para a análise de regressão, os três genótipos (AA, Aa e aa) foram codificados como 1, 0 e -1 para quantificação do efeito aditivo e como 0, 1 e 0 para quantificação do efeito atribuído à dominância, utilizando o modelo:

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

sendo,

Y_j é o valor da característica quantitativa avaliada no j-ésimo indivíduo da população F_2 ;

X_{1j} é o código do marcador para o efeito aditivo (AA = 1, Aa = 0 e aa = -1);

X_{2j} é o código do marcador para o efeito dominante (AA = 0, Aa = 1 e aa = 0);

β_0 é o intercepto (média da característica);

β_1 é o coeficiente de regressão associado ao efeito aditivo;

β_2 é o coeficiente de regressão associado ao efeito dominante; e

ε_j é o erro aleatório manifestado na característica no j-ésimo indivíduo.

2.3.2. Método de intervalo simples

O método de detecção de QTL por intervalo foi proposto por Lander; Botstein (1989) com a finalidade de testar a possibilidade de ocorrência do loco controlador da característica (QTL) de uma série de posições dentro do intervalo flanqueados por dois marcadores, utilizando a regressão linear para estimação dos parâmetros. Este modelo de regressão, para população F_2 , é dado por:

$$Y_j = \mu + ax_j^* + dz_j^* + \varepsilon_j$$

sendo,

Y_j é o valor da característica Y no indivíduo j;

μ é a média da característica da população;

a é o efeito aditivo do loco que está sendo estudado sobre a característica;

d é o efeito de dominância do loco que está sendo estudado sobre a característica;

x_j^* e z_j^* são as variáveis condicionadoras e dependentes dos genótipos dos marcadores que flanqueiam o QTL, no indivíduo j; e

ε_j é o erro aleatório.

2.3.3. Método de intervalo composto

No mapeamento de QTL, os testes que avaliam um intervalo de cada vez, sem considerar o restante da informação do genoma, podem apresentar problemas de detecção, principalmente se houver outros QTLs de grande efeito no mesmo grupo ou em outros grupos de ligação. Isso ocorre com o mapeamento por intervalo simples, cuja análise baseia-se exclusivamente nos marcadores que flanqueiam o intervalo avaliado (SCHUSTER; CRUZ, 2008).

A inclusão de marcadores que não estão associados com a característica pode levar a um aumento na variância residual, ao invés de redução. Portanto, o ideal é que se utilize apenas um subconjunto de marcadores selecionados como cofatores (c), que serão utilizados na análise. Os cofatores considerados nas análises foram selecionados utilizando-se o procedimento de regressão de *stepwise*, com $\alpha = 0,05$ para inclusão e exclusão dos marcadores no modelo (SCHUSTER; CRUZ, 2008; CRUZ et al., 2009).

Assim, para um melhor entendimento, utilizou-se o mapeamento de QTLs por intervalo composto (ZENG, 1993).

$$Y_j = [\mu + ax_j^* + dz_j^*] + \sum_{k \neq i, i+1}^c b_k x_{kj} + \varepsilon_j$$

sendo,

y_j é o valor da característica y no indivíduo j ;

x_{kj} é o escore do marcador k no indivíduo j (MM=1, Mm=0, mm=-1);

b_k é o coeficiente de regressão parcial do fenótipo Y com o k -ésimo marcador;

ε_j é o erro aleatório;

O termo entre colchetes é o mesmo apresentado no modelo de regressão para mapeamento de intervalo simples.

2.4. Eficiência do tamanho populacional sobre o mapeamento e obtenção de estimativas de efeitos de QTLs

Foram realizadas análises de variância, considerando o modelo fixo, das distâncias adjacentes levando em consideração os quatro tamanhos populacionais (200, 400, 600 e 1000 indivíduos) para cada grupo de ligação (GL1, GL 2, GL 3, GL 4 e GL 5). Cada distância do grupo de ligação foi considerada uma repetição e como são 50 repetições de cada grupo de ligação, totalizando 1000 repetições (20 distâncias em cada grupo e 50 repetições).

Realizaram-se também análises de variâncias, considerando o modelo fixo, dos efeitos das médias genóticas, aditivos, devidos ao desvio de dominância e contribuição (R^2) gênica, para o método de marca simples. Essa análise também foi realizada para os efeitos das médias genóticas, aditivos, devidos ao desvio de dominância, contribuição gênica (R^2) e posição do QTL para o método de intervalo. Tanto para o método da marca simples quanto para o de intervalo, as análises de variância foram realizadas para os grupos de ligação 1, 2, 3 e 4, pois no GL 5 é sabido que não há presença de QTL. Como os valores paramétricos dos efeitos são conhecidos, estas análises permitem avaliar a eficiência do tamanho população em estimar o efeito de cada QTL simulado.

Esses efeitos foram comparados entre si pelo teste Tukey a 5% de probabilidade. Posteriormente, foram realizados os testes t de Student entre as médias obtidas pelas análises de cada tamanho de população e as médias simuladas. O modelo do teste t utilizado, é dado por:

$$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

sendo,

\bar{x} é média dos efeitos;

μ_0 é o valor da constante utilizado para comparação com a média dos efeitos;
s é o desvio padrão amostral;
n é o tamanho da amostra.

As análises de variância e os testes de médias foram realizados por meio do aplicativo computacional GENES (CRUZ, 2006).

2.5. Eficiência do tamanho populacional na detecção de QTL

Primeiramente foram verificados quais os tamanhos de população recuperam os grupos de ligação e em que não ocorrem inversões das marcas. Posteriormente, foram realizados, para as metodologias de detecção de QTL por meio da marca simples, a análise das frequências dos marcadores que estão ligados ao QTL. Como a posição do QTL é conhecida, a população ideal é a que posicionar o QTL mais próximo da posição simulada.

Na metodologia de intervalo foi analisada a distribuição das repetições em relação a posição do QTL detectado em cada tamanho populacional. Sabendo que a localização correta do QTL simulado é a 50 cM, o tamanho de população mais eficiente será aquele que proporcionar um maior número de repetições com a posição do QTL mais próximo deste valor.

Além dessas análises gráficas, também foram realizadas análises de variância e teste de médias para os efeitos avaliados. O tamanho populacional mais eficiente é aquele que obtiver médias mais próximas às simuladas e com os menores valores de variância.

3. RESULTADO E DISCUSSÃO

3.1. Teste de segregação

As análises de segregação, para as marcas individuais, com base no teste de qui-quadrado (χ^2) referente à uma repetição da população F₂ de 200 indivíduos estão representadas na Tabela 2. Para todos os tamanhos populacionais e em todas as repetições, as marcas segregam na proporção de 1:2:1, como o esperado e sem distorção, indicando que os tamanhos populacionais utilizados são apropriados para constatação do princípio da segregação postulado na primeira Lei de Mendel.

Para realização dos testes de segregação foi utilizado o software GQMOL (CRUZ, 2008), usando o nível de significância estabelecido pela proteção de Bonferroni. A utilização da proteção de Bonferroni visa reduzir a probabilidade de se cometer o erro tipo I, mantendo a significância total do experimento α_T , igual a 5% (CRUZ et al, 2009). Esta correção é muito utilizada nos testes de segregação quando se utiliza um número grande de marcadores.

Etges et al. (2009) tiveram como material de estudo a *Drosophila majovensis* e, para a detecção de QTLs responsáveis à produção de hidricarboneto (composto químico constituído essencialmente por carbono e hidrocarboneto), avaliaram 1650 progênies e 21 marcadores microssatélites sendo o teste de segregação realizado com a proteção de Bonferroni. Palhares (2010) verificou a segregação mendeliana de 455 marcadores AFLP sem distorções usando teste de χ^2 ao nível de 5% de significância. Posteriormente, realizaram os testes de segregação, com nível de significância obtido considerando a proteção de Bonferroni, tendo como consequência a constatação de outras 145 marcas sem distorção de segregação, resultando em 600 marcas para a construção do mapa genético. Vários outros trabalhos relatam o uso da proteção de Bonferroni no estabelecimento do nível de significância para testes individuais em análise genômica (XU et al., 2011; YAN et al., 2009; CARLBORG et al., 2005; ALCOCHETE et al., 2005; FLINT-GARCIA et al., 2005).

Tabela 2. Teste de segregação (Hipótese de segregação igual a 1:2:1) dos 105 marcadores ao nível conjunto de 5% de probabilidade pelo teste χ^2 com proteção de Bonferroni.

Marcador	Classes			Qui-quadrado	Probabilidade(%)
	AA	Aa	AA		
C ₁₁	57	103	40	3,07	21,55 ^{ns}
C ₁₂	56	105	39	3,39	18,36 ^{ns}
C ₁₃	56	99	45	1,23	54,06 ^{ns}
C ₁₄	54	97	49	0,43	80,65 ^{ns}
C ₁₅	52	97	51	0,19	90,94 ^{ns}
C ₁₆	53	96	51	0,36	83,53 ^{ns}
C ₁₇	46	102	52	0,44	80,25 ^{ns}
C ₁₈	51	99	50	0,03	98,51 ^{ns}
C ₁₉	47	104	49	0,36	83,53 ^{ns}
C ₁₁₀	45	101	54	0,83	66,03 ^{ns}
C ₁₁₁	42	104	54	1,76	41,48 ^{ns}
C ₁₁₂	40	105	55	2,75	25,28 ^{ns}
C ₁₁₃	36	105	59	5,79	5,53 ^{ns}
C ₁₁₄	38	104	58	4,32	11,53 ^{ns}
C ₁₁₅	39	105	56	3,39	18,36 ^{ns}
C ₁₁₆	47	103	50	0,27	87,37 ^{ns}
C ₁₁₇	43	108	49	1,64	44,04 ^{ns}
C ₁₁₈	45	103	52	0,67	71,53 ^{ns}
C ₁₁₉	44	105	51	0,99	60,96 ^{ns}
C ₁₂₀	45	105	50	0,75	68,73 ^{ns}
C ₁₂₁	50	97	53	0,27	87,37 ^{ns}
C ₂₁	64	90	46	5,24	7,28 ^{ns}
C ₂₂	65	85	50	6,75	3,42 ^{ns}
C ₂₃	62	90	48	3,96	13,81 ^{ns}
C ₂₄	55	98	47	0,72	69,77 ^{ns}
C ₂₅	56	94	50	1,08	58,27 ^{ns}
C ₂₆	57	95	48	1,31	51,94 ^{ns}
C ₂₇	60	96	44	2,88	23,69 ^{ns}
C ₂₈	59	95	46	2,19	33,45 ^{ns}
C ₂₉	53	105	42	1,71	42,53 ^{ns}
C ₂₁₀	49	105	46	0,59	74,45 ^{ns}
C ₂₁₁	50	106	44	1,08	58,27 ^{ns}
C ₂₁₂	49	109	42	2,11	34,82 ^{ns}
C ₂₁₃	46	109	45	1,63	44,26 ^{ns}
C ₂₁₄	47	103	50	0,27	87,37 ^{ns}
C ₂₁₅	48	100	52	0,16	92,31 ^{ns}
C ₂₁₆	47	100	53	0,36	83,53 ^{ns}
C ₂₁₇	50	98	52	0,12	94,18 ^{ns}
C ₂₁₈	48	100	52	0,16	92,31 ^{ns}
C ₂₁₉	53	97	50	0,27	87,37 ^{ns}
C ₂₂₀	55	89	56	2,43	29,67 ^{ns}
C ₂₂₁	56	86	58	3,96	13,81 ^{ns}
C ₃₁	50	90	60	3,00	22,31 ^{ns}
C ₃₂	45	94	61	3,28	19,40 ^{ns}
C ₃₃	44	95	61	3,39	18,36 ^{ns}
C ₃₄	46	96	58	1,76	41,48 ^{ns}
C ₃₅	49	98	53	0,24	88,69 ^{ns}
C ₃₆	48	99	53	0,27	87,37 ^{ns}
C ₃₇	52	96	52	0,32	85,21 ^{ns}
C ₃₈	50	99	51	0,03	98,51 ^{ns}
C ₃₉	48	100	52	0,16	92,31 ^{ns}
C ₃₁₀	50	97	53	0,27	87,37 ^{ns}
C ₃₁₁	50	94	56	1,08	58,27 ^{ns}
C ₃₁₂	49	97	54	0,43	80,65 ^{ns}
C ₃₁₃	47	104	49	0,36	83,53 ^{ns}

Tabela 2. Continuação.

C ₃₁₄	51	98	51	0,08	96,08 ^{ns}
C ₃₁₅	54	96	50	0,48	78,66 ^{ns}
C ₃₁₆	50	97	53	0,27	87,37 ^{ns}
C ₃₁₇	46	103	51	0,43	80,65 ^{ns}
C ₃₁₈	47	103	50	0,27	87,37 ^{ns}
C ₃₁₉	47	103	50	0,27	87,37 ^{ns}
C ₃₂₀	47	106	47	0,72	69,77 ^{ns}
C ₃₂₁	45	107	48	1,07	58,57 ^{ns}
C ₄₁	41	94	65	6,48	3,92 ^{ns}
C ₄₂	51	85	64	6,19	4,53 ^{ns}
C ₄₃	49	89	62	4,11	12,81 ^{ns}
C ₄₄	47	96	57	1,32	51,68 ^{ns}
C ₄₅	45	102	53	0,72	69,77 ^{ns}
C ₄₆	51	97	52	0,19	90,94 ^{ns}
C ₄₇	50	93	57	1,47	47,95 ^{ns}
C ₄₈	49	92	59	2,28	31,98 ^{ns}
C ₄₉	47	90	63	4,56	10,23 ^{ns}
C ₄₁₀	49	94	57	1,36	50,66 ^{ns}
C ₄₁₁	48	95	57	1,31	51,94 ^{ns}
C ₄₁₂	50	96	54	0,48	78,66 ^{ns}
C ₄₁₃	48	101	51	0,11	94,65 ^{ns}
C ₄₁₄	52	95	53	0,51	77,49 ^{ns}
C ₄₁₅	54	98	48	0,44	80,25 ^{ns}
C ₄₁₆	57	102	41	2,64	26,71 ^{ns}
C ₄₁₇	56	100	44	1,44	48,67 ^{ns}
C ₄₁₈	55	97	48	0,67	71,53 ^{ns}
C ₄₁₉	55	101	44	1,23	54,06 ^{ns}
C ₄₂₀	58	98	44	2,04	36,06 ^{ns}
C ₄₂₁	58	90	52	2,36	30,73 ^{ns}
C ₅₁	51	96	53	0,36	83,53 ^{ns}
C ₅₂	50	99	51	0,03	98,51 ^{ns}
C ₅₃	50	100	50	0,00	100,00 ^{ns}
C ₅₄	51	96	53	0,36	83,53 ^{ns}
C ₅₅	55	99	46	0,83	66,03 ^{ns}
C ₅₆	57	96	47	1,32	51,68 ^{ns}
C ₅₇	60	93	47	2,67	26,32 ^{ns}
C ₅₈	57	98	45	1,52	46,77 ^{ns}
C ₅₉	56	92	52	1,44	48,67 ^{ns}
C ₅₁₀	56	95	49	0,99	60,96 ^{ns}
C ₅₁₁	57	98	45	1,52	46,77 ^{ns}
C ₅₁₂	55	102	43	1,52	46,77 ^{ns}
C ₅₁₃	54	108	38	3,84	14,66 ^{ns}
C ₅₁₄	54	109	37	4,51	10,49 ^{ns}
C ₅₁₅	48	112	40	3,52	17,20 ^{ns}
C ₅₁₆	51	103	46	0,43	80,65 ^{ns}
C ₅₁₇	51	103	46	0,43	80,65 ^{ns}
C ₅₁₈	52	98	50	0,12	94,18 ^{ns}
C ₅₁₉	54	97	49	0,43	80,65 ^{ns}
C ₅₂₀	55	95	50	0,75	68,73 ^{ns}
C ₅₂₁	52	99	49	0,11	94,65 ^{ns}

Nível crítico em teste individual igual a 0,0488388 %

3.2. Recuperação dos grupos de ligação

No mapeamento das populações F_2 , verificou-se que foram formados cinco grupos de ligação com 21 marcadores cada. O número de grupos de ligação recuperado esteve de acordo com o esperado, tendo em vista que o genoma original, estabelecido por simulação, continha cinco grupos de ligação. Após a formação destes grupos, verificou-se, pelo método SARF, a melhor ordem dos marcadores que constituem cada um dos grupos.

Constatou-se que os tamanhos das populações investigados (200, 400, 600 e 1000 indivíduos) foram adequados para recuperar o número de grupos de ligação e ordenar os marcadores como estabelecido no genoma original, pois não foram encontradas inversões das marcas (Figura 3).

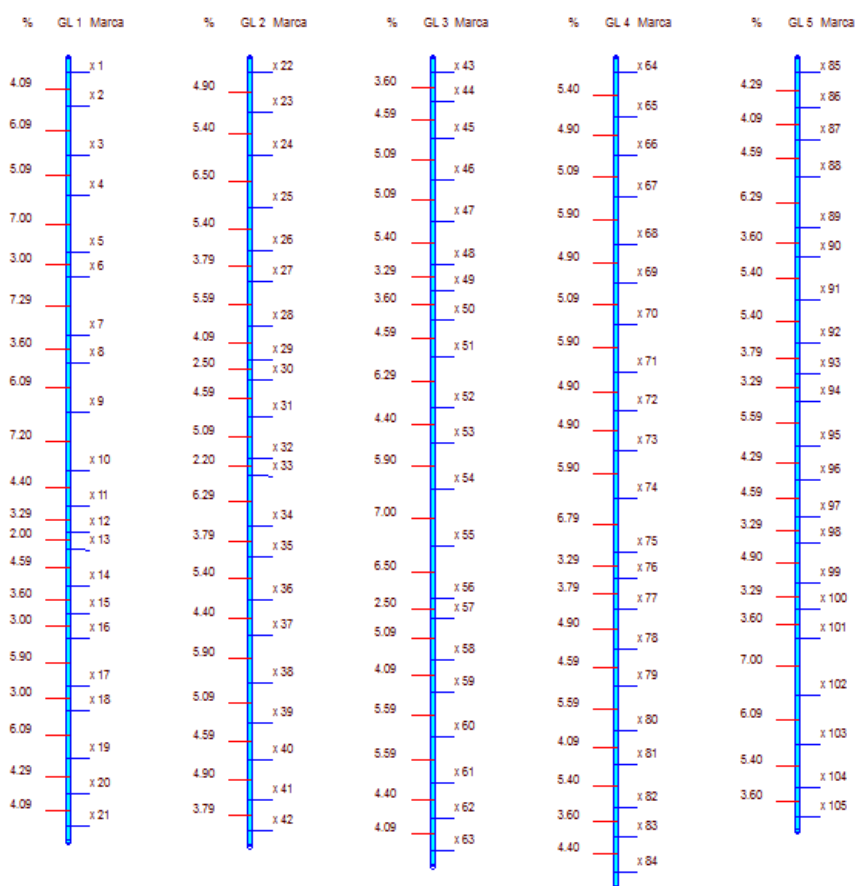


Figura 3. Mapa genético representativo da análise de uma população F_2 com tamanho populacional de 200 indivíduos.

Barros (2007), trabalhando com simulação em genotipagem seletiva, avaliou vários tamanhos (200, 300, 500 e 1000 indivíduos) de população F_2 e detectou que os

grupos de ligação de todos os tamanhos foram recuperados e as ordens das marcas foram as esperadas; resultados semelhantes foram obtidos por Cruz (2006).

Silva (2005) estudou vários tamanhos de populações RIL's (50, 100, 154, 200, 300, 500 e 800 indivíduos) para mapeamento. Neste trabalho verificou-se que quanto maior a população melhor é a recuperação dos grupos de ligação, tanto no tamanho quanto nas distâncias entre os marcadores. O autor constatou que os tamanhos mínimos de população para obtenção de mapas foram de 200, 300 e 500 indivíduos com alta saturação (5 cM), mediana (10 cM) e baixa (20 cM), respectivamente. Os tamanhos populacionais de 50 e 100 indivíduos proporcionaram mapas genéticos inadequados pela ocorrência de inversões no posicionamento entre marcas, em algumas repetições, para todos os níveis de saturação.

É importante ressaltar que muitos trabalhos citados na literatura têm por objetivo apenas avaliar a adequação do tamanho populacional para fins de mapeamento, visando estabelecer qual tamanho populacional melhor recupera os grupos e as informações dos marcadores (ordenamento e distâncias). Pouco tem sido estudado se o tamanho da população adequada para o mapeamento também é apropriado para a detecção, posicionamento dos QTLs e quantificação do efeito e importância de cada QTL.

3.3. Distância entre marcas adjacentes

Com o intuito de verificar se há diferenças entre as distâncias dos marcadores, em relação ao tamanho da população, foi realizada uma análise de variância considerando as repetições de simulação para cada grupo de ligação. A média das distâncias entre marcas adjacentes, ao longo de cada grupo de ligação, foi obtida fazendo-se duas médias aritméticas sucessivas. A primeira foi realizada entre os valores das distâncias entre as marcas dentro do mesmo grupo de ligação (20 distâncias). A segunda foi obtida das médias aritméticas anteriormente obtidas (50 repetições), com um total de 1000 repetições.

Na Tabela 3, encontra-se o resultado da análise de variância referente aos quatro tamanhos de população para os cinco grupos de ligação. Os efeitos de populações foram não significativos para todos os grupos de ligação, mostrando que as distâncias entre os marcadores adjacentes foram semelhantes nos quatro tamanhos de população. Pelo fato de que as populações foram simuladas, as distâncias médias entre as marcas adjacentes eram conhecidas. Assim sendo, esperava-se que as médias se aproximassem ao máximo das distâncias simuladas, que era 5 cM. Como pode ser observado nas Tabelas 3 e 4, os valores

foram muito próximos do valor simulado, não tendo diferenças entre as populações. Apesar disto, a população de 1000 indivíduos obteve os menores desvios padrão, mostrando maior precisão e confiabilidade nas distâncias.

Tabela 3. Resultado da análise de variância das distâncias entre as marcas adjacentes nos grupos de ligação 1 (GL 1), 2 (GL 2), 3 (GL 3), 4 (GL 4) e 5 (GL 5) das populações simuladas de 200, 400, 600 e 1000 indivíduos.

FV	GL	Quadrados Médios				
		GL 1	GL 2	GL 3	GL 4	GL 5
Repetições	999	0,6588	0,6681	0,5901	0,5925	0,6254
Populações	3	0,4999 ^{ns}	0,3806 ^{ns}	0,7673 ^{ns}	1,6715 ^{ns}	0,0400 ^{ns}
Resíduo	2997	0,6487	0,6647	0,6146	0,6450	0,6558
Média		4,99	4,97	4,95	4,97	4,96
CV(%)		16,13	16,40	15,85	16,17	16,32

^{ns} – não significativo pelo teste F.

Tabela 4. Médias e desvio padrão (entre parênteses) das distâncias entre as marcas adjacentes nos grupos de ligação 1 (GL 1), 2 (GL 2), 3 (GL 3), 4 (GL 4) e 5 (GL 5) das populações simuladas de 200, 400, 600 e 1000 indivíduos.

Tamanho	GL 1	GL 2	GL 3	GL 4	GL 5
200	4,98 (1,14)	4,98 (1,13)	4,97 (1,05)	4,91 (1,09)	4,97 (1,16)
400	5,02 (0,83)	4,97 (0,83)	4,92 (0,79)	5,01 (0,83)	4,95 (0,77)
600	4,99 (0,66)	4,99 (0,66)	4,93 (0,66)	4,96 (0,63)	4,96 (0,63)
1000	4,97 (0,50)	4,94 (0,51)	4,97 (0,50)	4,98 (0,51)	4,96 (0,49)

Silva (2005) avaliou as distâncias entre marcadores adjacentes de sete tamanhos de população RIL's (50, 100, 154, 200, 300, 500 e 800 indivíduos). Foi verificado que, em todos os grupos de ligação com saturação de 5 cM, as populações de 100 a 800 indivíduos não diferiram entre si. Tais resultados indicam que o comportamento da população F₂ é similar a outros tipos de populações.

3.4. Detecção de QTL

3.4.1. Detecção de QTL por marca simples baseada em regressão linear

Em 1975, Geldermann sugeriu a utilização do termo QTL para locos controladores de características quantitativas. Desde então, foi relatada a possibilidade de quantificar a associação entre o marcador e o QTL, por meio de um teste de comparação de médias ou

pela regressão dos valores da característica quantitativa, em função dos escores relativos ao genótipo do gene marcador. Neste caso, codifica-se os três genótipos para se efetuar a análise de regressão, sendo comumente utilizados os códigos 1, 0 e -1 para os efeitos aditivos e 0, 1 e 0 para os efeitos atribuídos à dominância, para os genótipos MM, Mm e mm, respectivamente (SCHUSTER; CRUZ, 2008).

3.4.1.1. Detecção de marcadores ligados a QTLs

Encontram-se, na Figura 4, a frequência dos marcadores, das 50 repetições, ligados aos QTLs dos respectivos grupos de ligação referentes às populações de tamanho 200, 400, 600 e 1000 indivíduos. Para a população 200, observa-se que, no grupo de ligação 1, há vários marcadores (9, 10, 11, 12, 13 e 14) ligados, com frequência de 100%, ao QTL. Resultado semelhante foram encontrados nos grupos de ligação 2 e 3, porém com menor números de marcadores (32 e 33, para o grupo de ligação 2; 53 para o grupo de ligação 3) ligados ao QTL em todas as repetições. Já para o grupo de ligação 4, o marcador 74 obteve a maior frequência, com 86%. Vale ressaltar que foram verificados em algumas repetições, no grupo de ligação 5, marcadores ligados ao QTL, com uma frequência inferior a 10%. Este resultado é um falso positivo, uma vez que, no processo de simulação, não foi alocado QTL neste grupo de ligação.

Observa-se na população de 400 indivíduos que, nos grupos de ligação 1, 2 e 3, os marcadores 7 a 17, 28 a 36 e 50 a 57, respectivamente, foram ligados ao QTL em todas as repetições. O marcador 74 foi o de maior frequência (98%) na detecção de QTL para o grupo de ligação 4. Resultados semelhantes foram obtidos para as populações de tamanho 600, no entanto, o marcador 74 foi significativo em todas as repetições. Para as populações de 1000 indivíduos a maioria dos marcadores referentes aos grupos de ligação 1, 2 e 3 foram significativos em todas as repetições e para o grupo 4 os marcadores 71 a 76 foram ligados ao QTL com uma frequência de 100%.

É importante ressaltar que cada um dos grupos de ligação 1, 2, 3 e 4 continham um QTL explicando 40%, 30%, 20% e 10%, respectivamente e o quinto grupo não continha QTL, sendo utilizado como controle da metodologia. A análise de regressão pelo método da marca simples, nesses grupos, possibilitou a detecção e obtenção de estimativas dos efeitos da média genotípica, aditivos, de dominância e a contribuição gênica (R^2), que expressa o quanto o QTL explica a variação genotípica da característica (SCHUSTER; CRUZ, 2008).

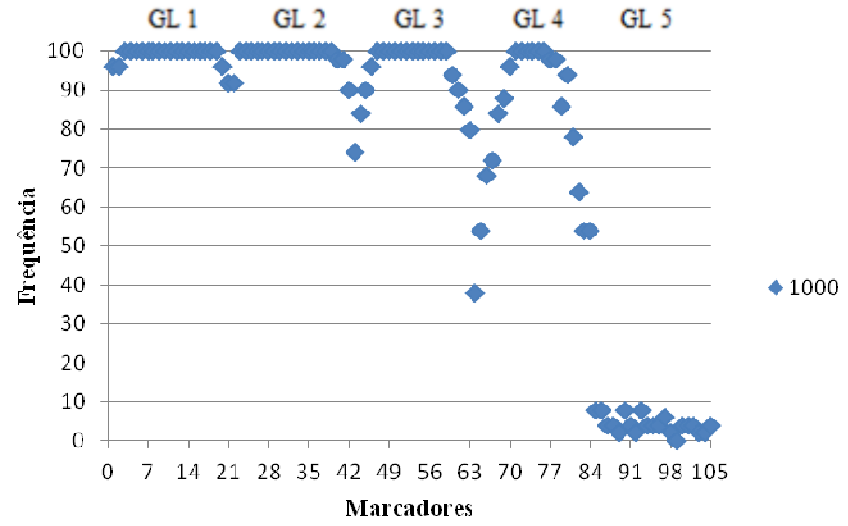
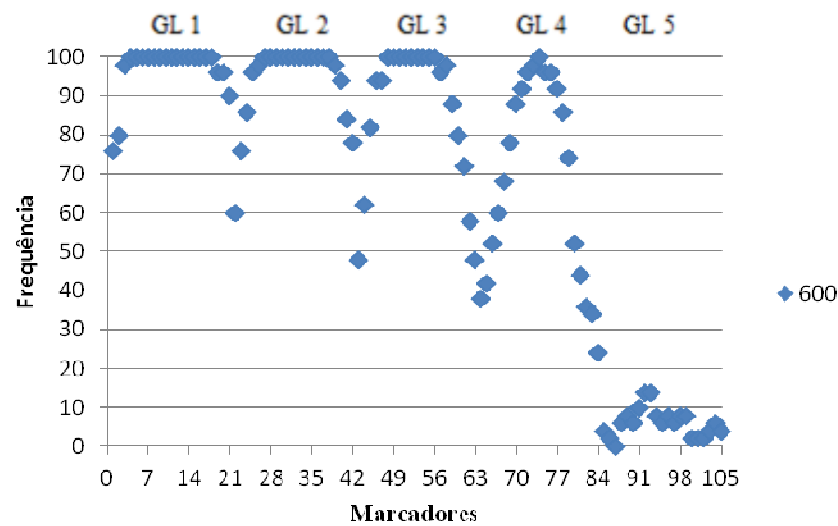
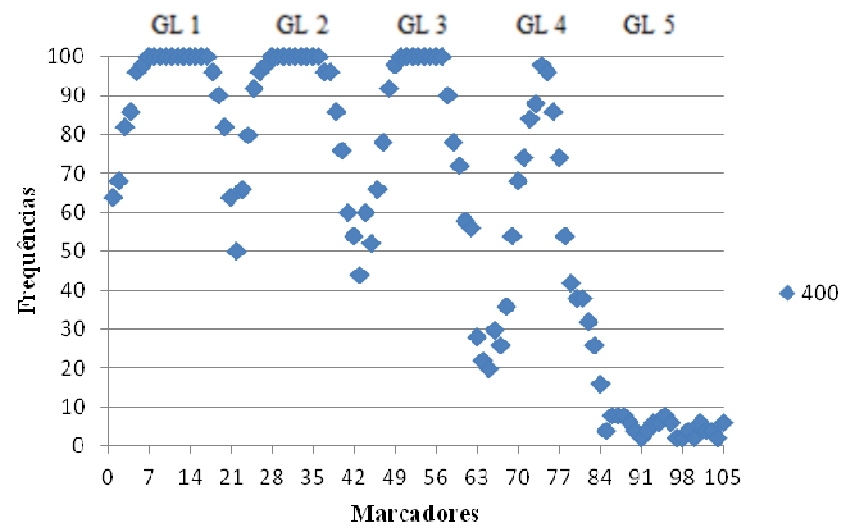
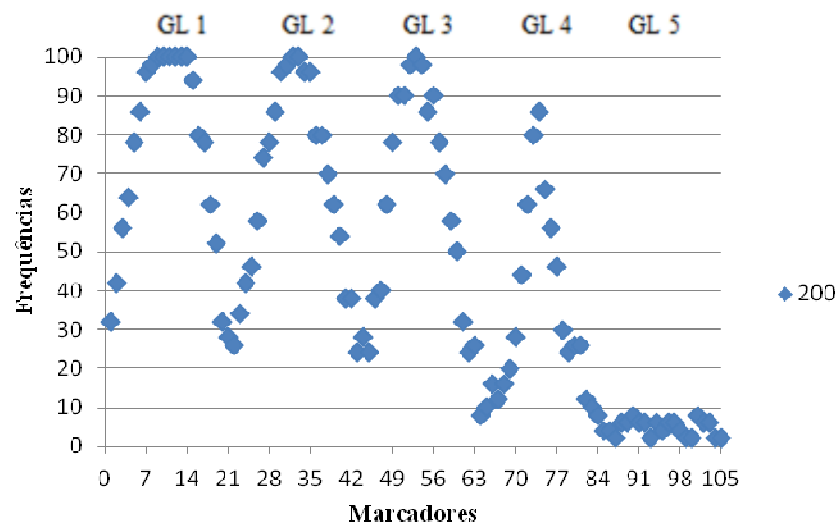


Figura 4. Frequências dos 105 marcadores ligados aos QTLs nos grupos de ligação 1, 2, 3, 4, e 5 obtidas nas populações simuladas de 200, 400, 600 e 1000 indivíduos. No eixo das ordenadas estão as frequências e no eixo das abscissas os marcadores moleculares.

Em relação ao grupo de ligação 1, praticamente não houve diferenças entre os tamanhos de população, pois em todos foram verificados marcadores ligados ao QTL. O maior valor de R^2 foi obtido, predominantemente, pelo marcador 11 em todas as populações e repetições.

No grupo de ligação 2 verificaram-se, na população de 200 indivíduos, dois marcadores associados ao QTL em 100% das repetições e nas demais populações verificaram-se vários marcadores com essa frequência. O marcador 32 foi o de maior R^2 , na maioria dos casos, em todos os tamanhos de população. Foi observado, no grupo de ligação 3, que apenas uma marca foi significativa em todas as populações de tamanho 200 e várias marcas para as demais populações. O marcador 53 obteve, na maioria das repetições, o maior R^2 para todos os tamanhos populacionais.

No grupo de ligação 4, o marcador 74 obteve frequências de 86% e 98% nas populações de 200 e 400 indivíduos, respectivamente, e 100% nas populações de 600; já nas de 1000, seis marcadores obtiveram esta frequência. No grupo de ligação 5, no qual não se esperava nenhum marcador ligado ao QTL, foram observados marcadores ligados com uma frequência máxima de 8% para as populações de 200, 400 e 1000 indivíduos e 14% para populações de 600.

Com o intuito de verificar associação entre marcas moleculares e QTL, Rodrigues et al (2010) avaliaram 207 famílias $F_{2:3}$ de soja para as características proteína e teor de óleo. Foram encontrados seis marcadores ligados a essas características e com elevado R^2 , indicando que esses QTLs são de grande efeito. Tais resultados são corroborados pelo presente trabalho, que verificou marcas ligadas aos QTLs que explicam de 20% a 40% da variação genotípica em todas as repetições. Detectaram-se também três marcas ligadas a QTLs de pequeno efeito, sendo duas explicando 10% e uma 4% da variação genotípica. De acordo com o presente trabalho, populações de 200 indivíduos não são consistentes na detecção de QTL que explica 10% da característica, portanto deve-se ter cautela em sua detecção.

3.4.1.2. Quantificação dos efeitos do QTL

Em cada grupo de ligação em que foi detectado o QTL, foi quantificado, na posição do marcador associado ao maior valor da razão de verossimilhança (LOD), os efeitos do QTL por meio do marcador mais próximo. Com o intuito de verificar se há diferenças entre

as estimativas dos efeitos (média genotípica, aditivo, de dominância e contribuição gênica) obtidos na população F₂, com diferentes tamanhos populacionais (200, 400, 600 e 1000 indivíduos), foi realizado a análise de variância.

Na análise de variância, referente ao grupo de ligação 1, observou-se que o tamanho populacional não proporcionou diferença significativa entre os efeitos dos QTLs quantificados na população F₂ (Tabela 5). Resultados semelhantes foram observados no grupo de ligação 2 (Tabelas 6). A análise de variância para este grupo, foi realizada levando em consideração dados perdidos, que são aqueles marcadores cujo efeito foi não significativo para efeitos aditivos ou de dominância.

Tabela 5. Análise de variância dos efeitos das médias (m), aditivos (a), devidos ao desvio de dominância (d) e da contribuição gênica (R²) das populações simuladas de 200, 400, 600 e 1000 indivíduos, referente ao grupo de ligação 1.

FV	Quadrados Médios				
	GL	m	A	d	R ²
Blocos	49	0,7437	0,6042	2,5646	9,2505
Populações	3	0,3522 ^{ns}	0,0516 ^{ns}	1,3906 ^{ns}	11,8023 ^{ns}
Resíduo	147	0,622	0,8357	1,7169	13,7426
Total	199				
Média		104,82	7,30	7,26	20,53
CV(%)		0,75	12,52	18,04	18,06

^{ns} – não significativo pelo teste F.

Tabela 6. Análise de variância dos efeitos das médias (m), aditivos (a), devidos ao desvio de dominância (d) e da contribuição gênica (R²) das populações simuladas de 200, 400, 600 e 1000 indivíduos, referente ao grupo de ligação 2.

FV	Quadrados Médios						
	GL	m	a	GL	d	GL	R ²
Blocos	49	0,625867	0,743647	49	1,344	49	5,55735
Populações	3	1,2561 ^{ns}	0,5222 ^{ns}	3	0,4573 ^{ns}	3	8,9171 ^{ns}
Resíduo	147	0,757479	0,920764	145	1,7064	147	10,37115
Total	199			197		199	
Média		105,27	6,43		6,36		15,78
CV(%)		0,83	14,93		20,55		20,41

^{ns} – não significativo pelo teste F.

No grupo de ligação 3, verificou-se diferenças significativas a 1% de probabilidade, entre os tamanho de populações para os efeitos das médias genotípica e contribuição gênica; significativos a 5% para os efeitos aditivos e de dominância (Tabela 7). A população de 200 indivíduos obteve a maior média genotípica, com 106,44 (Tabela 8). Em relação aos efeitos aditivos e de desvios de dominância, apesar de serem significativos pelo teste F, o teste

Tukey não detectou diferenças entre os tamanhos de populações. O maior valor da contribuição gênica foi obtido pela população de 400 indivíduos, com 11, 29; porém, não diferiu, estatisticamente, da população de 200 (10,89).

Tabela 7. Análise de variância dos efeitos das médias (m), aditivos (a), devidos ao desvio de dominância (d) e da contribuição gênica (R^2) das populações simuladas de 200, 400, 600 e 1000 indivíduos, referente ao grupo de ligação 3.

FV	Quadrados Médios							
	GL	m	GL	a	GL	d	GL	R^2
Blocos	49	0,980961	49	0,73814	49	1,46833	49	5,45458
Populações	3	4,2093**	3	2,5576*	3	2,8271*	3	34,6344**
Resíduo	147	0,642322	145	0,94317	131	0,96552	147	8,083187
Total	199		197		183		199	
Média		106,02		5,19		5,20		10,38
CV(%)		0,76		18,72		18,91		27,39

* e ** – significativo a 5% e 1%, respectivamente, pelo teste F.

Tabela 8. Médias dos efeitos das médias (m), aditivos (a), devidos ao desvio de dominância (d) e da contribuição gênica (R^2) das populações simuladas de 200, 400, 600 e 1000 indivíduos, referente ao grupo de ligação 3.

Tamanho	m	A	d	R^2
200	106,44 a	5,44 a	5,47 a	10,89 ab
400	105,77 b	5,33 a	5,41 a	11,29 a
600	105,89 b	4,98 a	5,02 a	9,69 b
1000	105,97 b	5,02 a	4,97 a	9,66 b

Médias seguidas da mesma letra na coluna não diferem entre si pelo teste Tukey a 5% de probabilidade.

Para os tamanhos de populações, no grupo de ligação 4, foram significativos a 1% para os efeitos aditivos e contribuição gênica, significativo a 5% para efeitos de dominância e não significativo para média genotípica (Tabela 9). A população de 200 indivíduos obteve os maiores valores de efeitos aditivos, de dominância e ação gênica, com valores de 4,15; 4,62 e 6,57, respectivamente (Tabela 10). No entanto, o valor do efeito de dominância não diferiu da população de tamanho 400 (4,20).

Tabela 9. Análise de variância dos efeitos das médias (m), aditivos (a), devidos ao desvio de dominância (d) e da contribuição gênica (R^2) das populações simuladas de 200, 400, 600 e 1000 indivíduos, referente ao grupo de ligação 4.

FV	Quadrados Médios							
	GL	M	GL	a	GL	D	GL	R^2
Blocos	49	1,236164	49	0,72162	49	2,27697	49	4,111443
Populações	3	0,832419 ^{ns}	3	4,3781**	3	4,4950*	3	23,7204**
Resíduo	146	0,848933	140	0,65221	104	1,29013	146	3,787468
Total	198		192		156		198	
Média		106,66		3,73		4,07		5,55
CV(%)		0,86		21,71		27,89		35,04

^{ns}, * e ** – não significativo, significativo a 5% e 1%, respectivamente, pelo teste F.

Tabela 10. Médias dos efeitos das médias (m), aditivos (a), devidos ao desvio de dominância (d) e da contribuição gênica (R^2) das populações simuladas de 200, 400, 600 e 1000 indivíduos, referente ao grupo de ligação 4.

Tamanho	m	A	d	R^2
200	106,85 a	4,15 a	4,62 a	6,57 a
400	106,62 a	3,39 b	4,20 ab	4,98 b
600	106,60 a	3,68 b	3,96 b	5,38 b
1000	106,57 a	3,71 b	3,75 b	5,31 b

Médias seguidas da mesma letra na coluna não diferem entre si pelo teste Tukey a 5% de probabilidade.

Observa-se que no grupo de ligação 1, onde o QTL explica 40% da característica, não houve diferença entre os tamanhos populacionais, o mesmo ocorreu com o grupo de ligação 2 (QTL explica 30%). Nos grupos de ligação 3 e 4, onde os QTLs explicam 20% e 10% da variação genotípica, respectivamente, houve diferenças para a maioria dos efeitos. Portanto, quanto menor a ação do QTL maiores serão as diferenças entre as populações e maior será o coeficiente de variação. Assim sendo, populações de tamanhos maiores serão mais precisos.

3.4.1.3. Comparação entre efeitos do QTL estimados e paramétricos

Nas Tabelas 11 e 12 estão apresentadas as médias e os valores do teste t referentes aos efeitos das médias genotípicas (m), aditivo (a), de dominância (d) e contribuição gênica referentes ao grupo de ligação 1. Observa-se que as médias e contribuição gênica foram significativas a 1% de probabilidade em todos os tamanhos populacionais, sendo que as médias foram superiores à simulada e as contribuições gênicas inferiores para todos os tamanhos. Os efeitos aditivos e de dominância não diferiram dos simulados, mostrando que os tamanhos de população obtiveram os

valores desses efeitos, similares aos simulados. Portanto, neste caso, pode-se utilizar população de tamanho 200 que irá obter valores de a e d próximos do real. Resultados semelhantes foram observados nos grupos de ligação 2 (Tabelas 13 e 14) e 3 (Tabelas 15 e 16).

Tabela 11. Médias, desvio padrão (entre parênteses) e valores do teste t para os efeitos das médias genotípicas (m) e aditivos (a) para as populações de 200, 400, 600 e 1000 indivíduos, referentes ao grupo de ligação 1.

Populações	m			A		
	Estimado	Simulada	Valor t	Estimado	Simulado	Valor t
200	104,77 (1,31)	100	25,79**	7,266 (1,25)	7,303	-0.21 ^{ns}
400	104,94 (0,57)	100	60,81**	7,290 (0,81)	7,303	-0.11 ^{ns}
600	104,81 (0,62)	100	54,46**	7,319 (0,71)	7,303	0.16 ^{ns}
1000	104,75 (0,41)	100	81,46**	7,339 (0,62)	7,303	0.41 ^{ns}

^{ns} e ** – não significativo e significativo 1%, respectivamente, pelo teste t.

Tabela 12. Médias, desvio padrão (entre parênteses) e valores do teste t para os efeitos devidos ao desvio de dominância (d) e contribuição gênica (R^2) para as populações de 200, 400, 600 e 1000 indivíduos, referentes ao grupo de ligação 1.

Populações	d			R^2		
	Estimado	Simulado	Valor t	Estimado	Simulado	Valor t
200	7,384 (2,13)	7,303	0,27 ^{ns}	21,00 (4,99)	40	-26,92**
400	7,016 (1,08)	7,303	-1,88 ^{ns}	19,91 (3,58)	40	-39,68**
600	7,311 (1,21)	7,303	0,05 ^{ns}	20,81 (2,87)	40	-47,30**
1000	7,339 (0,71)	7,303	0,36 ^{ns}	20,38 (2,12)	40	-65,55**

^{ns} e ** – não significativo e significativo 1%, respectivamente, pelo teste t.

Tabela 13. Médias, desvio padrão (entre parênteses) e valores do teste t para os efeitos da médias genotípicas (m) e aditivos (a) para as populações de 200, 400, 600 e 1000 indivíduos, referentes ao grupo de ligação 2.

Populações	m			A		
	Estimado	Simulado	Valor t	Estimado	Simulado	Valor t
200	105,51 (1,20)	100	32,47**	6,424 (1,33)	6,3246	0,53 ^{ns}
400	105,21 (0,93)	100	39,50**	6,493 (0,94)	6,3246	1,27 ^{ns}
600	105,21 (0,61)	100	60,57**	6,511 (0,74)	6,3246	1,78 ^{ns}
1000	105,16 (0,48)	100	76,08**	6,286 (0,54)	6,3246	-0,51 ^{ns}

^{ns} e ** – não significativo e significativo 1%, respectivamente, pelo teste t.

Tabela 14. Médias, desvio padrão (entre parênteses) e valores do teste t para os efeitos devidos ao desvio de dominância (d) e contribuição gênica (R^2) para as populações de 200, 400, 600 e 1000 indivíduos, referentes ao grupo de ligação 2.

Populações	d			R^2		
	Estimado	Simulado	Valor t	Estimado	Simulado	Valor t
200	6,205 (1,68)	6,3246	-0,50 ^{ns}	15,94 (4,30)	30	-23,11**
400	6,374 (1,36)	6,3246	0,26 ^{ns}	15,97 (2,91)	30	-34,09**
600	6,385 (1,00)	6,3246	0,42 ^{ns}	16,06 (2,54)	30	-38,72**
1000	6,453 (0,92)	6,3246	0,99 ^{ns}	15,15 (1,77)	30	-59,45**

^{ns} e ** – não significativo e significativo 1%, respectivamente, pelo teste t.

Tabela 15. Médias, desvio padrão (entre parênteses) e valores do teste t para os efeitos da médias genotípicas (m) e aditivos (a) para as populações de 200, 400, 600 e 1000 indivíduos, referentes ao grupo de ligação 3.

Populações	m			A		
	Estimado	Simulado	Valor t	Estimado	Simulado	Valor t
200	106,44 (1,36)	100	33,57**	5,421 (1,31)	5,164	1,39 ^{ns}
400	105,77 (0,75)	100	54,04**	5,335 (0,88)	5,164	1,38 ^{ns}
600	105,89 (0,51)	100	81,68**	4,984 (0,87)	5,164	-1,47 ^{ns}
1000	105,97 (0,49)	100	86,17**	5,021 (0,60)	5,164	-1,69 ^{ns}

^{ns} e ** – não significativo e significativo 1%, respectivamente, pelo teste t.

Tabela 16. Médias, desvio padrão (entre parênteses) e valores do teste t para os efeitos devidos ao desvio de dominância (d) e contribuição gênica (R^2) para as populações de 200, 400, 600 e 1000 indivíduos, referentes ao grupo de ligação 3.

Populações	d			R^2		
	Estimado	Simulado	Valor t	Estimado	Simulado	Valor t
200	5,463 (1,37)	5,164	1,54 ^{ns}	10,89 (3,89)	20	-16,56**
400	5,409 (1,17)	5,164	1,47 ^{ns}	11,29 (2,67)	20	-23,07**
600	5,025 (0,86)	5,164	-1,14 ^{ns}	9,69 (2,24)	20	-32,57**
1000	4,972 (0,81)	5,164	-1,67 ^{ns}	9,66 (1,56)	20	-47,00**

^{ns} e ** – não significativo e significativo 1%, respectivamente, pelo teste t.

No grupo de ligação 4, Tabelas 17 e 18, as médias genotípicas foram superiores às simuladas e as contribuições gênicas foram inferiores, em todas as populações. Em relação ao efeito aditivo as populações de tamanho 200 e 400 obtiveram diferenças significativas e as de 600 e 1000 indivíduos foram não significativas. As diferenças para os valores dos efeitos de dominância foram significativos para os tamanhos 200, 400 e 600 e não significativo para 1000 indivíduos, não diferindo dos valores simulados.

Tabela 17. Médias, desvio padrão (entre parênteses) e valores do teste t para os efeitos da médias genotípicas (m) e aditivos (a) para as populações de 200, 400, 600 e 1000 indivíduos, referentes ao grupo de ligação 4.

Populações	m			A		
	Estimado	Simulado	Valor t	Estimado	Simulado	Valor t
200	106,85 (1,42)	100	34,16**	4,175 (1,16)	3,6515	3,19**
400	106,62 (0,94)	100	49,90**	3,386 (0,75)	3,6515	-2,49*
600	106,6 (0,75)	100	61,81**	3,678 (0,71)	3,6515	0,27 ^{ns}
1000	106,56 (0,51)	100	90,97**	3,706 (0,59)	3,6515	0,65 ^{ns}

^{ns}, * e ** – não significativo, significativo a 5% e 1%, respectivamente, pelo teste t.

Tabela 18. Médias, desvio padrão (entre parênteses) e valores do teste t para os efeitos devidos ao desvio de dominância (d) e contribuição gênica (R^2) para as populações de 200, 400, 600 e 1000 indivíduos, referentes ao grupo de ligação 4.

Populações	d			R^2		
	Estimado	Simulado	Valor t	Estimado	Simulado	Valor t
200	4,582 (2,06)	3,6515	3,20**	6,56 (2,73)	10	-8,92**
400	4,133 (1,21)	3,6515	2,81**	4,98 (1,88)	10	-18,92**
600	4,065 (0,96)	3,6515	3,05**	5,37 (1,73)	10	-18,90**
1000	3,729 (0,85)	3,6515	0,64 ^{ns}	5,31 (1,26)	10	-26,30**

^{ns}, * e ** – não significativo, significativo a 5% e 1%, respectivamente, pelo teste t.

A população de 1000 indivíduos obteve os melhores resultados, com vários marcadores ligados ao QTL em 100% das repetições, no quarto grupo de ligação. Destes resultados aliados aos do teste t, pode-se inferir que nas populações de 1000 indivíduos os resultados obtidos são similares aos resultados de simulação em todas as situações estudadas.

Ding et al (2011) estudaram QTLs ligados à produtividade em milho e, para isto, utilizaram uma população F_2 de 225 indivíduos. Foram verificados QTLs de efeito aditivo, tanto com efeito positivo quanto negativo, variando de -6,5 a 7,9. Os QTLs detectados explicaram mais de 25% da variação genotípica; portanto, mesmo com uma população menor, os efeitos detectados são próximos aos reais.

3.4.2. Detecção de QTL por intervalo

3.4.2.1. Posição do QTL

Com o intuito de encontrar o tamanho populacional que melhor detecta e posiciona QTLs, sendo que estes explicaram diferentes porcentagens da variação

genotípica, foram realizadas análises estatísticas e gráficas com as posições dos QTLs, em cM. Estas posições, das 50 repetições, estão representadas nas Figuras 5, 6, 7 e 8 referentes aos grupos de ligação 1, 2, 3 e 4, respectivamente.

Observa-se que, no grupo de ligação 1, para a população de 1000 indivíduos obteve-se a menor variação. A maior amplitude foi obtida pela população de tamanho 200 (38,9 cM a 57,7 cM); no entanto, as variações foram similares. Esta mesma população obteve também, no grupo de ligação 2, a maior variação, sendo que as posições variaram de 23,4 cM a 59 cM e os demais tamanhos obtiveram variações menores e similares. Resultados semelhantes foram observados no grupo de ligação 3, que explica 20% da variação genotípica.

A população de tamanho 1000, no grupo de ligação 4, foi a que menos variou (46,3 cM a 53,1 cM), seguida pela população de tamanho 600 (29,4 cM a 59 cM), 400 (31,4 cM a 78,1 cM) e 200 (0 cM a 86,8 cM).

A metodologia de mapeamento por intervalo foi desenvolvida por Lander; Botstein (1989) com o intuito de franquear o QTL por meio de duas marcas, permitindo posicionar o gene no mapa de ligação. Com o objetivo de encontrar genótipos de mandioca mais produtivos, Boonchanawiwat et al. (2011) utilizaram uma população F_2 de 100 indivíduos e por meio de intervalo simples encontraram vários QTLs de grande e médio efeitos ligados à produtividade. Esta metodologia é amplamente encontrada na literatura (CHAKRABORTY; ZENG, 2011; BOROVSKY; PARAN, 2011; MCCLOSKEY et al., 2011).

As metodologias de intervalo simples são eficientes para detectar e posicionar o QTL. No entanto, há situações em que esta metodologia não é eficiente, detectando QTL fantasma. Para solucionar este problema pode-se utilizar a metodologia de intervalo composto, que se destaca por levar em consideração os cofatores, que são marcadores do mesmo ou de outros grupos de ligação. Desta forma, consegue-se minimizar a interferência do efeito de outros QTLs, tornando a detecção do QTL de interesse mais precisa (LYNCH; WALSH, 1998; WU et al, 2007; SCHUSTER; CRUZ, 2008). Moro (2011) utilizou uma população F_2 de milho com 512 indivíduos para detecção de QTLs. Para isto, foi utilizado a metodologia de intervalo composto, sendo que os marcadores utilizados como cofatores foram selecionados por meio de regressão de *stepwise*. Foram verificados, na literatura, vários trabalhos com intervalo composto para detecção de QTL (LIU et al., 2011; DING et al., 2011; LAPERUTA, 2011; LU et al, 2011; YUAN et al, 2011).

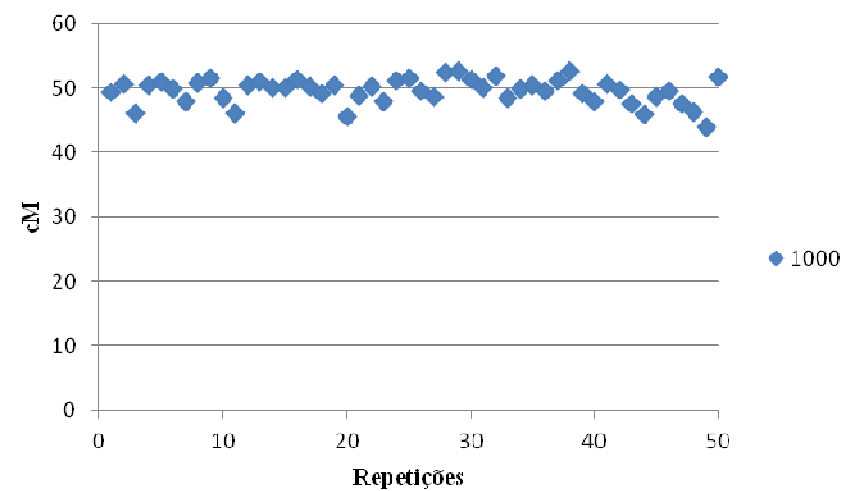
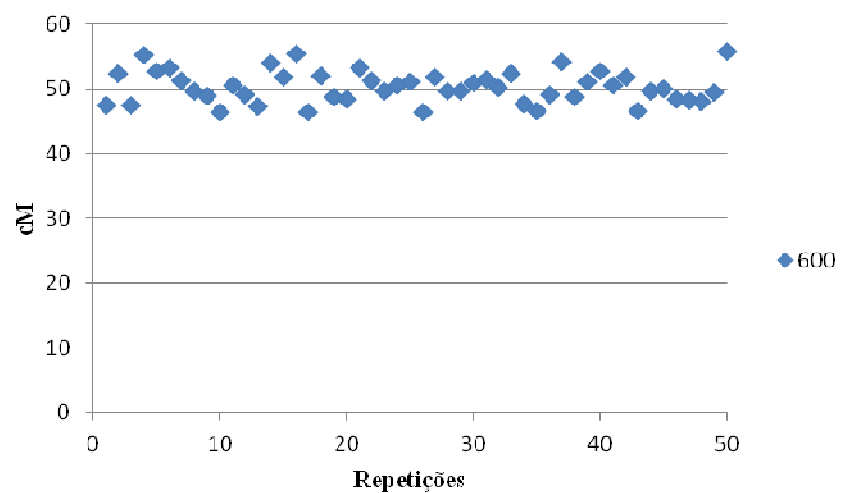
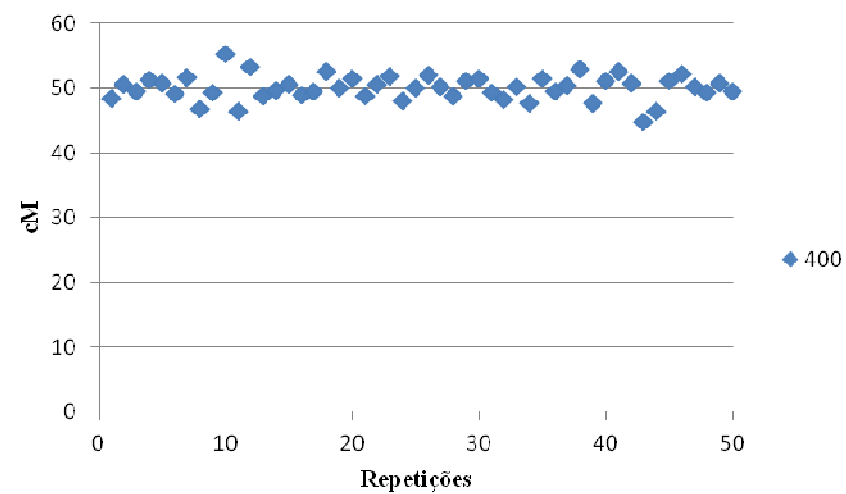
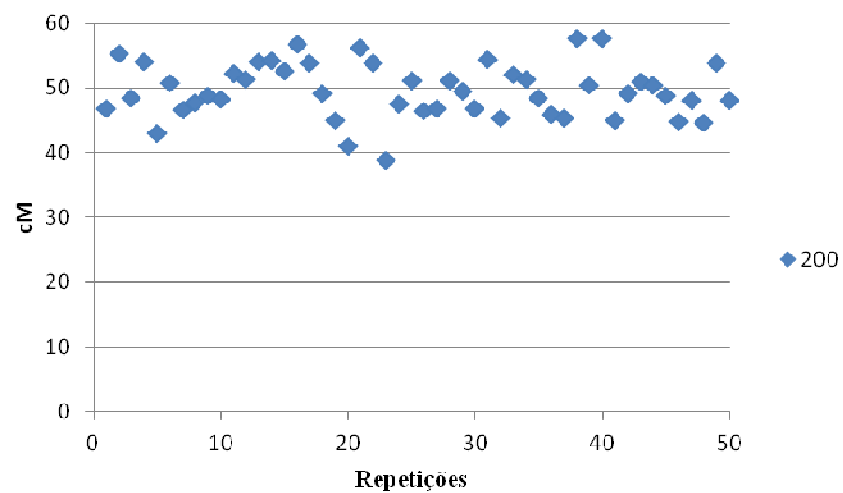


Figura 5. Distribuição das posições dos QTLs das 50 repetições obtidas nas populações simuladas de 200, 400, 600 e 1000 indivíduos referentes ao grupo de ligação 1. No eixo das ordenadas estão as posições, em cM, e no eixo das abscissas as repetições.

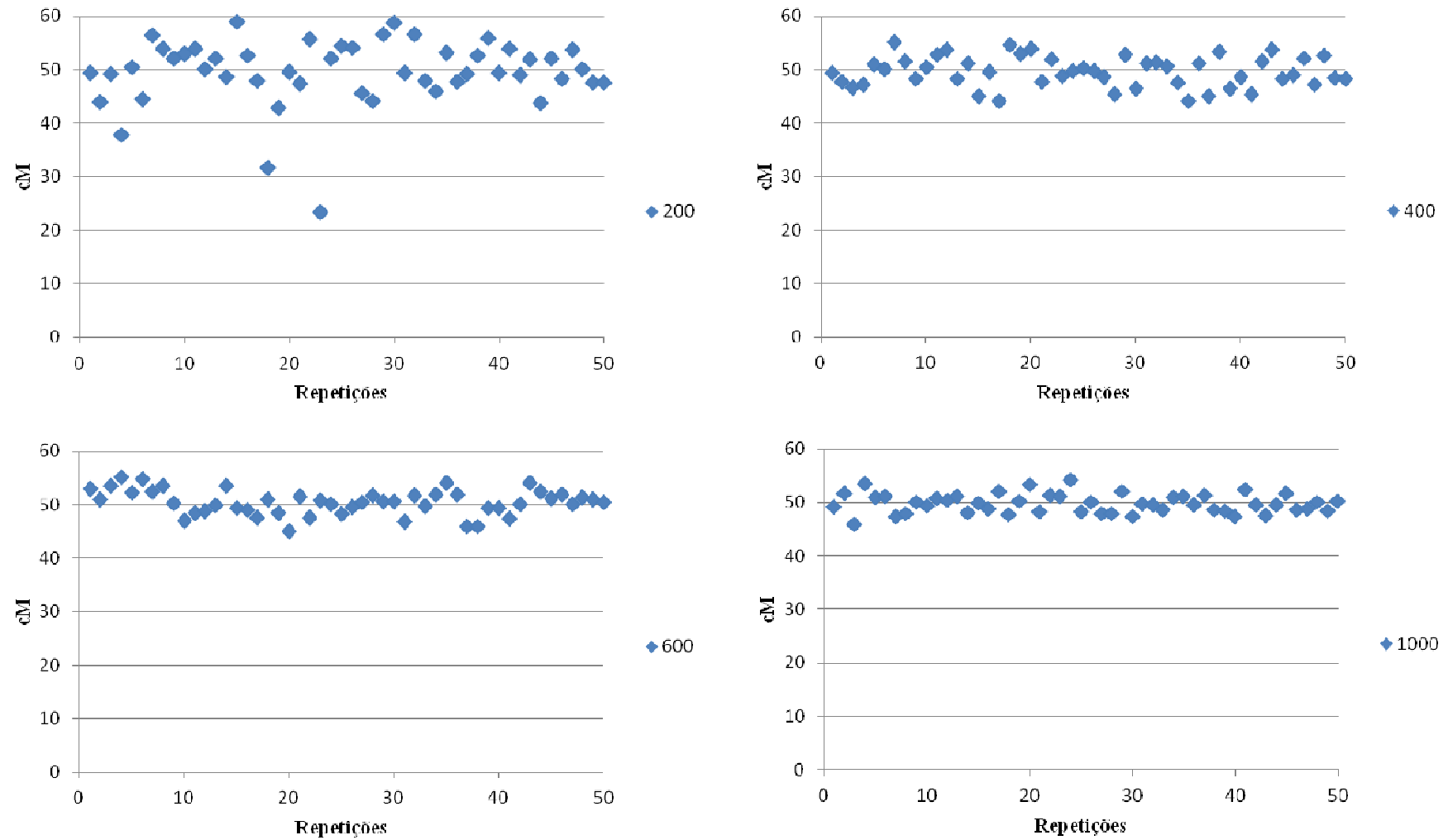


Figura 6. Distribuição das posições dos QTLs das 50 repetições obtidas nas populações simuladas de 200, 400, 600 e 1000 indivíduos referentes ao grupo de ligação 2. No eixo das ordenadas estão as posições, em cM, e no eixo das abscissas as repetições.

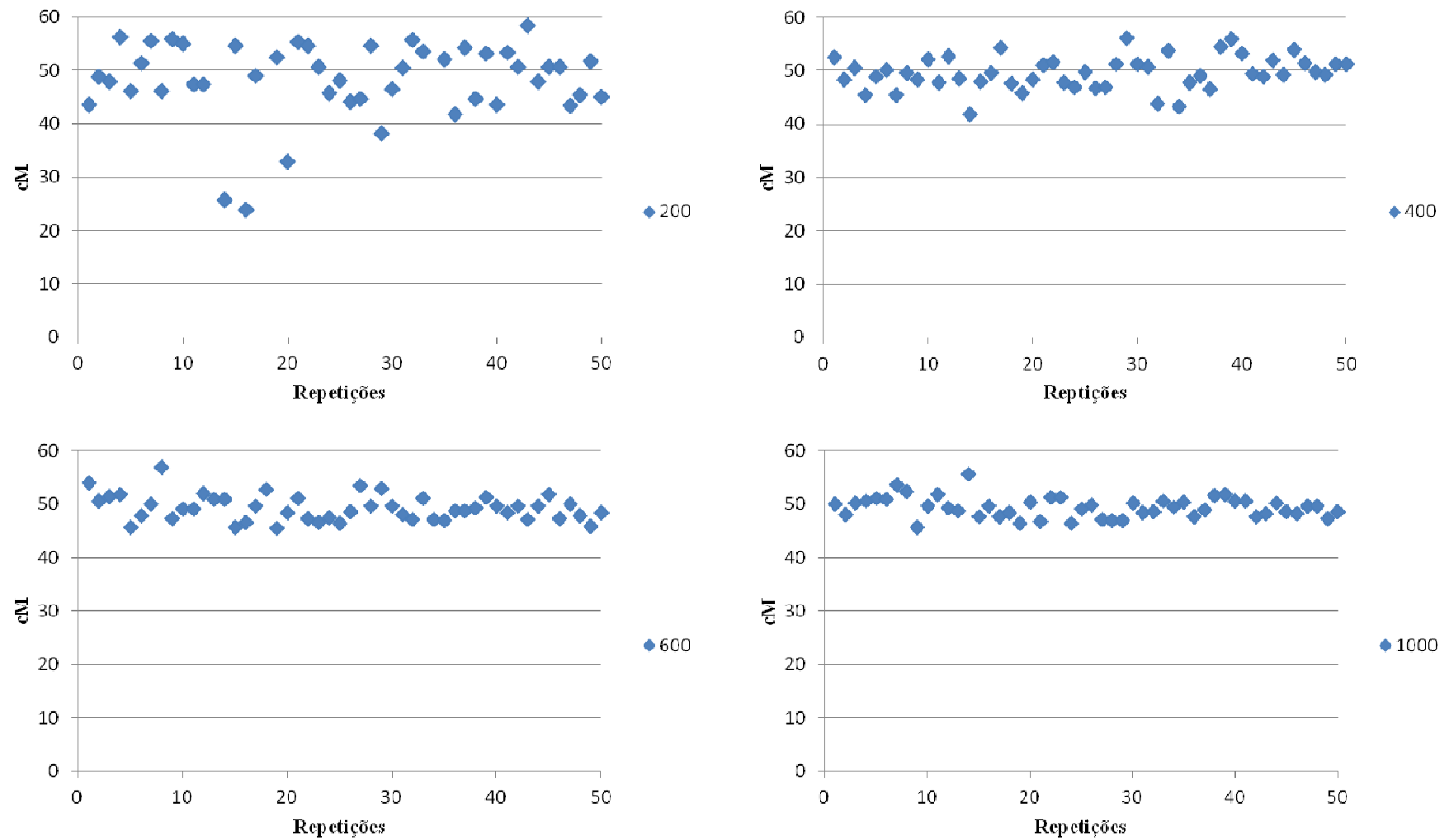


Figura 7. Distribuição das posições dos QTLs das 50 repetições obtidas nas populações simuladas de 200, 400, 600 e 1000 indivíduos referentes ao grupo de ligação 3. No eixo das ordenadas estão as posições, em cM, e no eixo das abscissas as repetições.

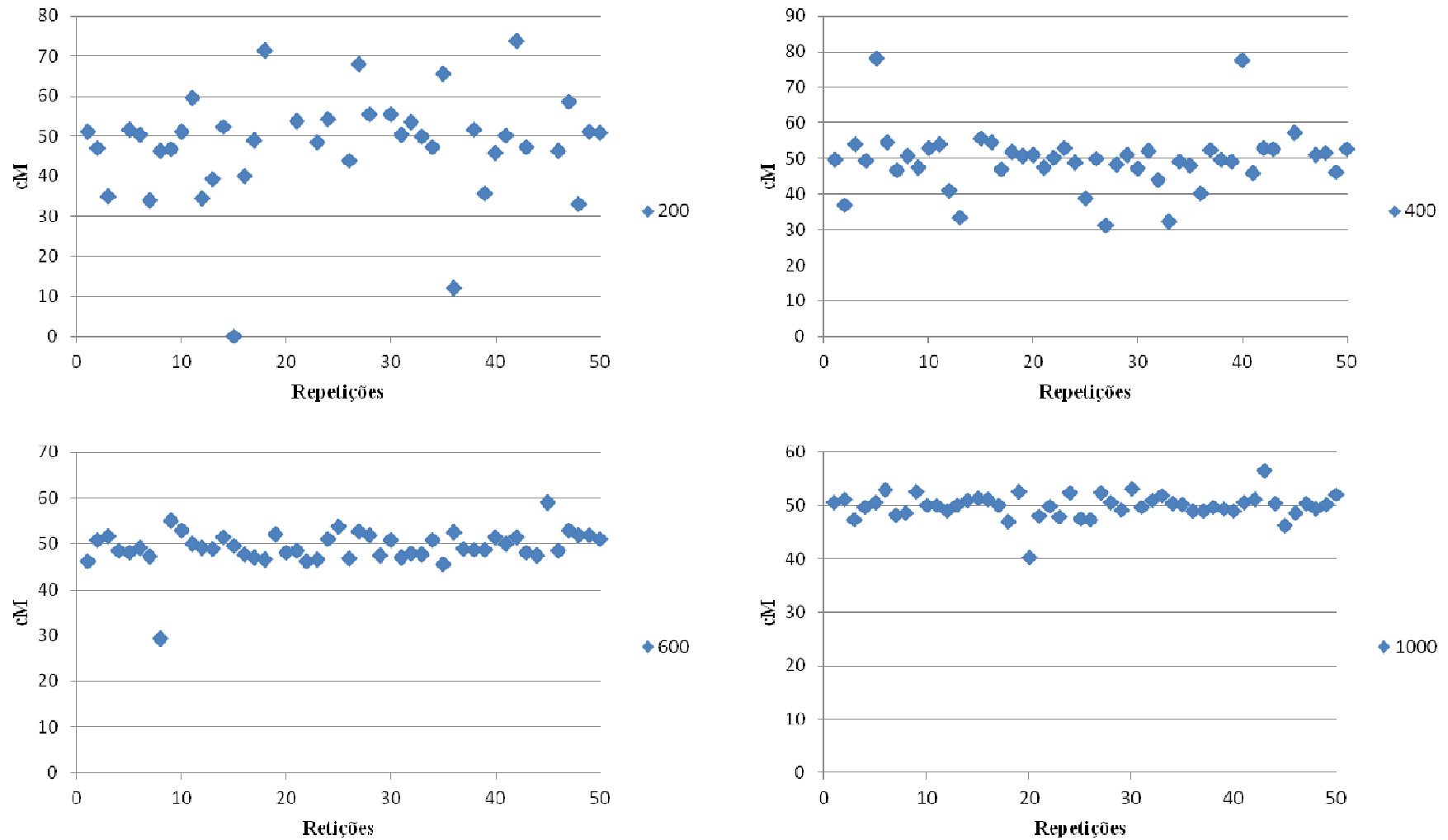


Figura 8. Distribuição das posições dos QTLs das 50 repetições obtidas nas populações simuladas de 200, 400, 600 e 1000 indivíduos referentes ao grupo de ligação 4. No eixo das ordenadas estão as posições, em cM, e no eixo das abscissas as repetições.

Observou-se que no grupo de ligação 1 não houve diferenças entre os tamanhos populacionais e nos grupos de ligação 2 e 3 os tamanhos de 400, 600 e 1000 indivíduos foram similares. Já no quarto grupo de ligação, as populações de 1000 indivíduos obtiveram os melhores resultados. Estes comportamentos, nos diferentes grupos de ligação, são devidos à contribuição de cada QTL na manifestação genotípica.

Na Tabela 19 encontram-se as médias e os desvios padrão das posições dos QTLs nos quatro grupos de ligação. Nas médias das posições, houve poucas diferenças entre as populações, variando de 47,82 cM (GL 4, 200 indivíduos) a 50,42 cM (GL 2, 600 indivíduos). Foi observado que quanto menos o QTL explica a característica, maior foi o valor dos desvios padrão. Pois, a população de tamanho 200 foi a que obteve os maiores valores de desvios padrão (4,20 a 13,38) e a de 1000 obteve os menores valores.

Tabela 19. Médias e desvio padrão (entre parênteses) das posições dos QTLs das populações de tamanho 200, 400, 600 e 1000 indivíduos referentes aos grupos de ligação 1, 2, 3 e 4.

Populações	GL 1	GL 2	GL 3	GL 4
200	49,64 (4,20)	49,51 (6,31)	48,27 (7,20)	47,82 (13,38)
400	50,13 (1,93)	49,61 (2,85)	50,00 (3,09)	49,55 (8,45)
600	50,39 (2,47)	50,42 (2,36)	50,00 (2,41)	50,00 (3,89)
1000	49,55 (1,94)	49,82 (1,79)	50,00 (1,95)	50,00 (2,29)

Yan et al. (2009) estudaram diferentes tamanhos de populações F_2 e herdabilidades. Para isto, foram simulados três tamanhos de populações (50, 100 e 200 indivíduos) e três diferentes herdabilidades (5%, 10% e 20%). O tamanho dos cromossomos simulado foi de 100 cM, sendo que a posição do QTL foi a 55 cM. A população de 200 foi a que menos variou, nos vários cenários, de 53,96 a 56,33 e 1,55 a 10,98 para posição e desvio padrão, respectivamente.

Estão apresentados na Tabela 20 os números de repetições que não detectaram QTLs, quando se deveria. É importante salientar que o número de repetições, utilizado neste trabalho, foi 50 e que nos grupos de ligação 1, 2, 3 e 4 os QTLs explicavam 40%, 30%, 20% e 10% da característica genotípica, respectivamente. Foi observado que, quanto menor a contribuição gênica do QTL, maior foi a porcentagem de repetições que não detectaram QTL. No entanto, este problema pode ser minimizado utilizando-se a metodologia de intervalo composto. Vale ressaltar que, mesmo com tal uso, pode ocorrer a não detecção do QTL.

Tabela 20. Número de repetições que não detectaram QTLs pelas metodologias de mapeamentos por intervalo simples (IS) e intervalo composto (IC) nos grupos de ligação 1, 2, 3 e 4 das populações simuladas de 200, 400, 600 e 1000 indivíduos.

GL	200 ind.		400 ind.		600 ind.		1000 ind.	
	IS	IC	IS	IC	IS	IC	IS	IC
1	18 (36) ¹	4 (8)	11 (22)	0	3 (6)	0	1 (2)	0
2	20 (40)	1 (2)	11 (22)	0	5 (10)	0	2 (4)	0
3	20 (40)	2 (4)	17 (34)	0	17 (34)	3 (6)	8 (16)	0
4	30 (60)	11 (22)	31 (62)	7 (14)	26 (52)	1 (2)	18 (36)	0

¹ – Números entre parênteses correspondem à percentagem.

Outro fator que influencia a detecção do QTL é o tamanho da população. A população de tamanho 200, mesmo utilizando a metodologia de intervalo composto, obteve repetições que não detectaram QTL. Nota-se que, quando utilizou-se o intervalo composto, houve uma diminuição da não detecção de QTL, conforme aumentou o tamanho populacional, sendo que na população de 1000 indivíduos todos os QTLs foram detectados. Portanto, quanto maior o tamanho populacional melhor será sua detecção.

Na Tabela 21 estão apresentadas as repetições que detectaram mais de um QTL por grupo de ligação. Verifica-se que quanto menor a ação do QTL, maior será o número de repetições que detectam mais de um QTL por grupo de ligação. Este problema também pode ser minimizado com a utilização da metodologia de intervalo composto. Entretanto, mesmo com esta metodologia o problema poderá persistir, caso se use tamanho populacional pequeno. Observa-se que quanto maior a população, menor será o número de repetições que detectam mais de um QTL por grupo, pois na população de tamanho 1000 nenhuma repetição detectou mais de um QTL e na de 600 este problema foi solucionado com o intervalo composto.

Tabela 21. Número de repetições que detectaram mais de 1 QTL por grupo de ligação pelas metodologias de mapeamentos por intervalo simples (IS) e intervalo composto (IC), nos grupos de ligação 1, 2, 3 e 4 das populações simuladas de 200, 400, 600 e 1000 indivíduos.

GL	200 ind.		400 ind.		600 ind.		1000 ind.	
	IS	IC	IS	IC	IS	IC	IS	IC
1	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0
3	3 (6) ¹	0	0	0	0	0	0	0
4	25 (50)	9 (18)	13 (26)	3 (6)	2 (4)	0	0	0

¹ – Números entre parênteses correspondem à percentagem.

No grupo de ligação 5, Tabela 22, observou-se uma frequência de quatro por cento na detecção de QTL, onde não deveria. Este problema foi solucionado utilizando a metodologia de intervalo composto, exceto para as populações de 400 indivíduos.

Tabela 22. Número de repetições que detectaram QTLs pelas metodologias de mapeamentos por intervalo simples (IS) e intervalo composto (IC), no grupo de ligação 5 das populações simuladas de 200, 400, 600 e 1000 indivíduos.

GL	200 ind.		400 ind.		600 ind.		1000 ind.	
	IS	IC	IS	IC	IS	IC	IS	IC
5	2 (4) ¹	0	2 (4)	2 (4)	1 (2)	0	2 (4)	0

¹ – Números entre parênteses correspondem à percentagem.

Do exposto, pode-se inferir que quanto menos o QTL explica a característica, maior deverá ser o tamanho da população de mapeamento. No grupo de ligação 1 praticamente não houve diferenças entre as populações. Já no grupo de ligação 4 houve grandes diferenças no posicionamento dos QTLs, sendo que o tamanho 1000 foi o mais preciso e adequado quando se deseja encontrar e posicionar QTLs que explicam pouco a característica.

São encontrados na literatura trabalhos que utilizam intervalo composto para dar maior precisão na detecção. Com o intuito de mapear os QTLs que controlam o florescimento do arroz em condições de seca, Chakraorty; Zeng (2011) utilizaram uma população de 154 duplos haplóides. Para a detecção dos QTLs foram utilizadas as metodologias de intervalo simples, composto e intervalos múltiplos, utilizando um valor crítico de LOD igual a 2,5. Com o intervalo simples foram detectados 12 QTLs, mas quando foram aplicadas as metodologias de intervalo composto e múltiplos foram detectados apenas cinco QTLs, em ambas as metodologias.

3.4.2.2. Efeitos dos QTLs

Foram realizadas análises de variância para os efeitos das médias genotípicas, aditivos, de dominância, posição e contribuição gênica dos QTLs das populações simuladas de tamanhos 200, 400, 600 e 1000 indivíduos. Estas análises foram realizadas para os grupos de ligação 1, 2, 3 e 4. Verificou-se que não houve diferenças entre os tamanhos populacionais para todos os efeitos, referentes ao grupo de ligação 1 e 2 (Tabelas 23 e 24).

Tabela 23. Análise de variância dos efeitos das médias (m), aditivos (a), devidos ao desvio de dominância (d), posição dos QTLs e contribuição gênica (R^2) das populações simuladas de 200, 400, 600 e 1000 indivíduos referente ao grupo de ligação 1.

FV	Quadrados Médios					
	GL	m	A	d	posição	R^2
Repetição	49	0,7066	0,6716	2,4963	9,8528	48,0016
Populações	3	0,5608 ^{ns}	0,0152 ^{ns}	2,2585 ^{ns}	8,0709 ^{ns}	118,3438 ^{ns}
Resíduo	147	0,6128	0,859	1,7228	7,1164	66,9444
Total	199					
Média		104,73	7,38	7,43	49,93	38,18
CV(%)		0,75	12,56	17,67	5,34	27,11

^{ns} – não significativo, pelo teste F.

Tabela 24. Análise de variância dos efeitos das médias (m), aditivos (a), devidos ao desvio de dominância (d), posição dos QTLs e contribuição gênica (R^2) das populações simuladas de 200, 400, 600 e 1000 indivíduos referente ao grupo de ligação 2.

FV	Quadrados Médios					
	GL	m	A	d	posição	R^2
Repetições	49	0,6677	0,7889	1,6161	10,9841	18,9228
Populações	3	1,3186 ^{ns}	0,5653 ^{ns}	3,2463 ^{ns}	8,3875 ^{ns}	21,3464 ^{ns}
Resíduo	147	0,8354	0,9464	2,3045	15,2389	32,9099
Total	199					
Média		105,21	6,50	6,42	49,84	20,77
CV(%)		0,87	14,97	23,64	7,83	27,62

^{ns} – não significativo, pelo teste F.

No grupo de ligação 3 os efeitos de tratamento foi não significativo para todos os efeitos, exceto para aqueles de dominância e contribuição gênica (Tabela 25). A população de tamanho 400 obteve o maior efeito de dominância (5,57). No entanto, não diferiram das populações de tamanho 600 e 1000. Os maiores valores de contribuição gênica foram obtidos pelas populações de 200 e 400 indivíduos (Tabela 26). É importante ressaltar que, neste caso, o esperado era que se aproximasse de 20, valor simulado.

Foram verificados que, no grupo de ligação 4, as diferenças entre os tamanhos populacionais foram significativas para os efeitos das médias genotípicas, aditivos e contribuição gênica (Tabela 27). A população de tamanho 1000 foi a que obteve a maior média (106,5); entretanto, não diferiu da população de 600 indivíduos e a população de tamanho 200 obteve as maiores estimativas do efeito aditivo e contribuição gênica (4,28 e 9,52, respectivamente), como pode ser observado na Tabela 28.

Tabela 25. Análise de variância dos efeitos das médias (m), aditivos (a), devidos ao desvio de dominância (d), posição dos QTLs e contribuição gênica (R^2) das populações simuladas de 200, 400, 600 e 1000 indivíduos referente ao grupo de ligação 3.

FV	Quadrados Médios					
	GL	m	A	d	posição	R^2
Repetição	49	1,67	0,9292	2,1599	15,8658	28,1337
Populações	3	0,5694 ns	1,3974 ns	6,8465 **	17,2094 ^{ns}	81,0873**
Resíduo	147	1,7325	1,0457	1,7067	17,3958	20,1103
Média		105,78	5,21	5,13	49,16	12,73
CV(%)		1,24	19,63	25,45	8,48	35,32

^{ns} e ** – não significativo e significativo a 1%, respectivamente, pelo teste F.

Tabela 26. Médias dos efeitos das médias (m), aditivos (a), devidos ao desvio de dominância (d), posição dos QTLs e contribuição gênica (R^2) das populações simuladas de 200, 400, 600 e 1000 indivíduos referente ao grupo de ligação 3.

Populações	m	a	d	posição	R^2
200	105,70 a	5,33 a	4,68 b	48,29 a	13,81 a
400	105,70 a	5,38 a	5,57 a	49,54 a	13,85 a
600	105,80 a	5,05 a	5,21 ab	49,34 a	11,57 b
1000	105,93 a	5,08 a	5,07 ab	49,47 a	11,69 b

Médias seguidas da mesma letra na coluna não diferem entre si pelo teste Tukey a 5% de probabilidade.

Tabela 27. Análise de variância dos efeitos das médias (m), aditivos (a), devidos ao desvio de dominância (d), posição dos QTLs e contribuição gênica (R^2) das populações simuladas de 200, 400, 600 e 1000 indivíduos referente ao grupo de ligação 4.

FV	Quadrados Médios					
	GL	M	a	d	posição	R^2
Repetições	49	13,394042	1,067765	2,532461	59,40307	13,7614
Populações	3	139,1458**	3,8593**	2,3802 ^{ns}	37,7067 ^{ns}	120,8074**
Resíduo	135	9,738276	0,766178	1,995324	63,201019	11,397
Total	187					
Média		105,08	3,80	3,88	49,24	6,90
CV(%)		2,97	23,02	36,38	16,15	48,90

^{ns}, * e ** – não significativo, significativo a 5% e 1%, respectivamente, pelo teste F.

Tabela 28. Médias dos efeitos das médias (m), aditivos (a), devidos ao desvio de dominância (d), posição dos QTLs e contribuição gênica (R^2) das populações simuladas de 200, 400, 600 e 1000 indivíduos referente ao grupo de ligação 4.

Populações	m	a	d	posição	R^2
200	102,50 c	4,28 a	4,23 a	47,80 a	9,52 a
400	104,80 b	3,57 b	3,87 a	49,56 a	6,16 b
600	106,16 ab	3,72 b	3,67 a	49,35 a	6,49 b
1000	106,50 a	3,76 b	3,83 a	49,98 a	5,85 b

Médias seguidas da mesma letra na coluna não diferem entre si pelo teste Tukey a 5% de probabilidade.

3.4.2.3. Comparação entre efeitos do QTL estimados e paramétricos

Com o intuito de verificar se há diferenças entre as estimativas obtidas e as simuladas, realizaram-se os testes t. Os valores deste teste para os grupos de ligação 1, 2, 3 e 4 estão apresentados nas Tabelas 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39 e 40. Para todas as populações as médias genotípicas foram maiores do que as simuladas em todos os grupos de ligação. Os efeitos aditivos foram não significativos para todas as populações nos grupos de ligação 1, 2 e 3, exceto para a população de tamanho 600, no grupo de ligação 2, que foi maior que o simulado. No grupo de ligação 4 a estimativa do efeito aditivo da população de 200 indivíduos foi superior ao valor simulado e as demais populações não diferiram.

Os efeitos devidos ao desvio de dominância foram não significativos para todas as populações nos grupos de ligação 1, 2 e 3, exceto para a população de 400 indivíduos, no grupo de ligação 3, que foi maior que o valor simulado. No grupo de ligação 4 a população de tamanho 200 obteve efeito superior aos valores simulados. Os valores das posições foram não significativos para todos os tamanhos de populações em todos os grupos de ligação. Já os valores das contribuições gênicas foram menores em todos os tamanhos e grupos de ligação, exceto para a população de 200 indivíduos no grupo 4, que foi não significativo.

Tabela 29. Médias, desvio padrão (entre parênteses) e valores do teste t para os efeitos das médias genotípicas (m) e aditivos (a) para as populações de 200, 400, 600 e 1000 indivíduos referentes ao grupo de ligação 1.

Tamanho	m			a		
	Estimado	Simulado	Valor t	Estimado	Simulado	Valor t
200	104,61 (1,28)	100	25,45**	7,3692 (1,28)	7,303	0,36 ^{ns}
400	104,87 (0,55)	100	62,87**	7,3606 (0,81)	7,303	0,51 ^{ns}
600	104,74 (0,64)	100	52,34**	7,3744 (0,73)	7,303	0,69 ^{ns}
1000	104,70 (0,44)	100	76,24**	7,401 (0,63)	7,303	1,10 ^{ns}

^{ns} e ** – não significativo e significativo a 1%, respectivamente, pelo teste t.

Tabela 30. Médias, desvio padrão (entre parênteses) e valores do teste t para os efeitos devidos ao desvio de dominância (d) e posição do QTL (posição) para as populações de 200, 400, 600 e 1000 indivíduos referentes ao grupo de ligação 1.

Tamanho	d			posição		
	Estimado	Simulado	Valor t	Estimado	Simulado	Valor t
200	7,676 (2,09)	7,303	1,26 ^{ns}	49,63 (4,20)	50	-0,62 ^{ns}
400	7,157 (1,06)	7,303	-0,97 ^{ns}	50,13 (1,93)	50	0,48 ^{ns}
600	7,456 (1,25)	7,303	0,86 ^{ns}	50,39 (2,43)	50	1,12 ^{ns}
1000	7,427 (0,76)	7,303	1,15 ^{ns}	49,54 (1,94)	50	-1,68 ^{ns}

^{ns} – não significativo pelo teste t.

Tabela 31. Médias, desvio padrão (entre parênteses) e valores do teste t para os efeitos da contribuição gênica (R^2) para as populações de 200, 400, 600 e 1000 indivíduos referentes ao grupo de ligação 1.

Tamanho	R^2		
	Estimado	Simulado	Valor t
200	32,20 (11,89)	40	-4,64**
400	28,66 (7,15)	40	-11,22**
600	30,46 (6,15)	40	-10,97**
1000	29,40 (4,36)	40	-17,19**

^{ns} – não significativo pelo teste t.

Tabela 32. Médias, desvio padrão (entre parênteses) e valores do teste t para os efeitos das médias genótípicas (m) e aditivos (a) para as populações de 200, 400, 600 e 1000 indivíduos referentes ao grupo de ligação 2.

Tamanho	M			a		
	Estimado	Simulado	Valor t	Estimado	Simulado	Valor t
200	105,45 (1,31)	100	29,30**	6,5188 (1,33)	6,3246	1,03 ^{ns}
400	105,14 (0,92)	100	39,66**	6,5496 (0,97)	6,3246	1,63 ^{ns}
600	105,13 (0,61)	100	59,64**	6,577 (0,74)	6,3246	2,41*
1000	104,95 (0,48)	100	72,98**	6,3412 (0,59)	6,3246	0,20 ^{ns}

^{ns}, * e ** – não significativo, significativo a 5% e 1%, respectivamente, pelo teste t.

Tabela 33. Médias, desvio padrão (entre parênteses) e valores do teste t para os efeitos devidos ao desvio de dominância (d) e posição do QTL (posição) para as populações de 200, 400, 600 e 1000 indivíduos referentes ao grupo de ligação 2.

Tamanho	d			posição		
	Estimado	Simulado	Valor t	Estimado	Simulado	Valor t
200	6,0406 (2,21)	6,3246	-0,91 ^{ns}	49,51 (6,31)	50	-0,55 ^{ns}
400	6,5258 (1,32)	6,3246	1,08 ^{ns}	49,61 (2,85)	50	-0,97 ^{ns}
600	6,5404 (1,03)	6,3246	1,48 ^{ns}	50,42 (2,36)	50	1,26 ^{ns}
1000	6,5786 (0,90)	6,3246	1,98 ^{ns}	49,82 (1,79)	50	-0,71 ^{ns}

^{ns}, * e ** – não significativo e significativo a 5%, respectivamente, pelo teste F.

Tabela 34. Médias, desvio padrão (entre parênteses) e valores do teste t para o efeito da contribuição gênica (R^2) para as populações de 200, 400, 600 e 1000 indivíduos referentes ao grupo de ligação 2.

Tamanho	R^2		
	Estimado	Simulado	Valor t
200	21,16 (7,54)	30	-8,29**
400	21,11 (5,42)	30	-11,61**
600	21,00 (4,74)	30	-13,43**
1000	19,79 (3,01)	30	-23,97**

^{ns} – não significativo pelo teste t.

Tabela 35. Médias, desvio padrão (entre parênteses) e valores do teste t para os efeitos das médias genóticas (m) e aditivos (a) para as populações de 200, 400, 600 e 1000 indivíduos referentes ao grupo de ligação 3.

Tamanho	m			a		
	Estimado	Simulado	Valor t	Estimado	Simulado	Valor t
200	106,19 (1,33)	100	32,90**	5,429 (1,39)	5,164	1.35 ^{ns}
400	105,7 (0,80)	100	50,38**	5,375 (0,94)	5,164	1.58 ^{ns}
600	105,79 (0,52)	100	78,79**	5,054 (0,87)	5,164	-0.90 ^{ns}
1000	105,92 (0,49)	100	85,45**	5,078 (0,61)	5,164	-1.00 ^{ns}

^{ns} e ** – não significativo e significativo a 1%, respectivamente, pelo teste t.

Tabela 36. Médias, desvio padrão (entre parênteses) e valores do teste t para os efeitos devidos ao desvio de dominância (d) e posição do QTL (posição) para as populações de 200, 400, 600 e 1000 indivíduos referentes ao grupo de ligação 3.

Tamanho	d			posição		
	Estimado	Simulado	Valor t	Estimado	Simulado	Valor t
200	4,75 (2,04)	5,164	-1,45 ^{ns}	48,26 (7,20)	50	-1,71 ^{ns}
400	5,571 (1,32)	5,164	2,18*	49,53 (3,09)	50	-1,07 ^{ns}
600	5,214 (0,90)	5,164	0,39 ^{ns}	49,33 (2,41)	50	-1,96 ^{ns}
1000	5,07 (0,83)	5,164	-0,80 ^{ns}	49,46 (1,95)	50	-1,95 ^{ns}

^{ns} – não significativo pelo teste t.

Tabela 37. Médias, desvio padrão (entre parênteses) e valores do teste t para o efeito da contribuição gênica (R^2) para as populações de 200, 400, 600 e 1000 indivíduos referentes ao grupo de ligação 3.

Tamanho	R^2		
	Estimado	Simulado	Valor t
200	13,81 (6,84)	20	-6,40**
400	13,86 (3,95)	20	-11,00**
600	11, 57 (2,94)	20	-20,27**
1000	11,69 (4,18)	20	-14,07**

^{ns} – não significativo pelo teste t.

Tabela 38. Médias, desvio padrão (entre parênteses) e valores do teste t para os efeitos das médias genóticas (m) e aditivos (a) para as populações de 200, 400, 600 e 1000 indivíduos referentes ao grupo de ligação 4.

Tamanho	M			a		
	Estimado	Simulado	Valor t	Estimado	Simulado	Valor t
200	102,46 (5,09)	100	3,42**	4,2386 (1,53)	3,6515	2,71**
400	104,71 (3,86)	100	8,62**	3,5475 (0,63)	3,6515	-1,16 ^{ns}
600	106,16 (2,18)	100	19,96**	3,7224 (0,70)	3,6515	0,71 ^{ns}
1000	106,49 (0,50)	100	91,78**	3,7618 (0,62)	3,6515	1,27 ^{ns}

^{ns}, * e ** – não significativo, significativo a 5% e 1%, respectivamente, pelo teste t.

Tabela 39. Médias, desvio padrão (entre parênteses) e valores do teste t para os efeitos devidos ao desvio de dominância (d) e posição do QTL (posição) para as populações de 200, 400, 600 e 1000 indivíduos referentes ao grupo de ligação 4.

Tamanho	d			posição		
	Estimado	Simulado	Valor t	Estimado	Simulado	Valor t
200	4,2079 (1,78)	3,6515	2,21*	47,82 (13,38)	50	-1,15 ^{ns}
400	3,8811 (1,65)	3,6515	0,98 ^{ns}	49,55 (8,45)	50	-0,38 ^{ns}
600	3,6717 (1,43)	3,6515	0,10 ^{ns}	49,35 (3,89)	50	-1,18 ^{ns}
1000	3,827 (0,90)	3,6515	1,38 ^{ns}	49,97 (2,29)	50	-0,09 ^{ns}

^{ns} e ** – não significativo e significativo a 5%, respectivamente, pelo teste t.

Tabela 40. Médias, desvio padrão (entre parênteses) e valores do teste t para o efeito da contribuição gênica (R^2) para as populações de 200, 400, 600 e 1000 indivíduos referentes ao grupo de ligação 4.

Tamanho	R^2		
	Estimado	Simulado	Valor t
200	9,48 (5,46)	10	-0,67 ^{ns}
400	6,16 (1,73)	10	-15,65**
600	6,48 (3,92)	10	-6,35**
1000	5,85 (1,51)	10	-19,43**

^{ns} – não significativo pelo teste t.

É importante salientar que o teste t leva em consideração os desvios padrão, sendo que quanto menor, maior será o valor t e, conseqüentemente, maior a probabilidade de ser significativo. Observam-se que os valores destes efeitos aditivos obtidos pelas populações de 200, 400 e 600 indivíduos, no grupo de ligação 2, foram 6,5188; 5,5496 e 6,577, com desvios padrão de 1,33; 0,97 e 0,74 e com os valores t de 1,03; 1,63 e 2,41, respectivamente. Os valores foram semelhantes, porém, os valores t foram discrepantes, sendo altamente influenciados pelos valores de desvio padrão. Outro exemplo encontra-se na Tabela 39, onde as posições dos QTLs, em média, dos tamanhos populacionais 200, 400, 600 e 1000 indivíduos, no grupos de ligação 4,

foram: 47,82 cM, 49,55 cM, 49,35 cM e 49,97 cM, com desvios padrão de 13,38; 8,45; 3,89 e 2,29 e os valores t de -1,15; -0,38; -1,18 e -0,09, respectivamente. Observa-se que o valor t da população de 600 indivíduos foi maior que o valor da população de tamanho 200. Entretanto, a posição da população de 600 foi mais próxima à posição simulada (50 cM) do que a de 200 indivíduos. Este fato pode ser explicado devido ao desvio padrão da população de tamanho 200 ser maior do que a de 600 indivíduos.

Vale ressaltar que, no geral, a população de tamanho 1000 obteve valores semelhantes aos simulados, pois foram não significativos exceto em relação às médias genotípicas e contribuição gênica. Outro fato relevante é que as variâncias da população 1000 foram menores, mostrando uma maior precisão nas estimativas.

Verificaram-se, na literatura, que a maioria dos trabalhos utiliza tamanhos de população próximos a 200 indivíduos (COUTO et al., 2010; CHAKRAORTY; ZENG, 2011; LU et al., 2011; BOROVSKY; PARAN, 2011). Entretanto, com o advento do sequenciamento são encontradas populações de tamanhos maiores, o que tornou as estimativas dos efeitos gênicos mais precisas (SILVA et al., 2011; SOUSA et al., 2011; TORTEREAU et al., 2011).

4.4.3. Considerações referentes à detecção de QTLs nas populações F₂

Em trabalhos de detecção de QTL, primeiramente precisa-se saber qual o objetivo da pesquisa: detecção ou detecção e posicionamento. Quando o objetivo for apenas a detecção, a metodologia da marca simples é eficiente. O tamanho populacional é de extrema importância nestes trabalhos, sendo que o melhor tamanho depende do quanto o QTL explica a característica. Assim, a população de 200 indivíduos é eficiente para detectar QTLs que expliquem 20% ou mais da característica, caso contrário necessita de uma população igual ou superior a 600 indivíduos. Quando o objetivo for detectar e posicionar, é recomendável a utilização das metodologias de intervalo. O tamanho populacional influencia diretamente na identificação da posição correta, sendo que quanto maior for a população, mais preciso será. Outro fato relevante é o quanto cada QTL explica a variação genotípica, pois quanto menos explica menor será a sua precisão e, conseqüentemente, maior deverá ser a população.

Quando o QTL explicou 40% da característica, os desvios padrão das populações de tamanho 400, 600 e 1000 foram semelhantes. Já quando explicou 30%, as populações de 600 e 1000 indivíduos foram similares. Resultados semelhantes a este foram observados no

grupo de ligação 3 (QTL explica 20%); no entanto, em 6% das populações de 600 indivíduos não foram detectados QTLs. O tamanho populacional de 1000 foi o mais preciso no posicionamento do QTL, quando este explicou apenas 10%. Portanto, no primeiro caso pode-se utilizar uma população de 400, no segundo 600, no terceiro e quarto de 1000 indivíduos.

As posições do QTL foram analisadas por meio das análises de variâncias e teste t e não foram detectadas diferenças entre os tamanhos populacionais nem entre os valores estimados e os simulados, todos nos grupos de ligação 1, 2, 3 e 4. Na média, os posicionamentos foram próximos aos simulados, no entanto, conforme aumentou o tamanho da população menores foram os desvios padrão.

Nas populações de 200, 400, 600 e 1000 indivíduos as estimativas dos efeitos aditivos e de dominância foram similares aos simulados onde os QTLs explicaram 40%, 30% e 20% da característica. Quando o QTL explicou 10% da característica a população 1000 obteve os melhores resultados. Entretanto, quando se utilizou a metodologia de intervalo, somente a de tamanho 200 obteve resultados diferentes dos simulados.

As estimativas das contribuições gênicas dos QTLs, tanto nas metodologias da marca simples quanto na de intervalo, foram subestimadas, em todas as populações. Este fato era esperado, uma vez que os efeitos foram medidos por meio de marcadores flanqueados e não no próprio QTL. No entanto, as estimativas obtidas pelas metodologias de intervalo, mesmo com as diferenças significativas, foram mais próximas às simuladas.

Encontram-se trabalhos que utiliza metodologia da marca simples (SANTOS et al, 2007; MIGNON et al, 2009), intervalo simples (MCCLOSKEY et al, 2011; SILVA et al, 2011; BOROVSKY; PARAN, 2011) e intervalo composto (LIU et al, 2011; YUAN et al, 2011; LU et al, 2011; DING et al, 2011). Também são encontrados trabalhos que utilizam o método da marca simples e os de intervalo para melhor entendimento (RODRIGUES et al, 2010).

4. CONCLUSÕES

Para fins de mapeamentos, conclui-se que:

As populações de 200, 400, 600 e 1000 indivíduos são eficientes para recuperar os grupos de ligação;

Em todos os tamanhos populacionais utilizados, as sequências das marcas foram como o esperado, sem ocorrência de inversões;

As distâncias entre as marcas adjacentes foram semelhantes nos quatro tamanhos de populações.

Para fins de detecção de QTL, conclui-se que:

Populações de 200 indivíduos são eficientes para detectar QTL que explica, no mínimo, 20% da variação genotípica.

Para QTL que explica de 10% a 20% as populações de 600 indivíduos são eficientes.

As populações de 200, 400, 600 e 1000 indivíduos são eficientes para detectar e posicionar QTLs que explicam acima de 40%, 30%, 20% e 10% da variação genotípica da característica, respectivamente.

Para fins de efeitos do QTL, conclui-se que:

Para os efeitos das médias genotípicas, os quatro tamanhos populacionais obtiveram valores superiores aos simulados;

As populações de 200, 400, 600 e 1000 indivíduos são eficientes para estimar os efeitos aditivos e de dominância quando o QTL explica 20% ou mais da característica;

Para QTLs que explicam entre 10% e 20% da variação genotípica a população de 1000 indivíduos é eficiente para estimar os efeitos aditivos e de dominância;

Os quatro tamanhos de população são eficientes em posicionar o QTL no grupo de ligação;

As populações de 200, 400, 600 e 1000 indivíduos subestimaram os efeitos da contribuição gênica do QTL;

Quanto maior o tamanho populacional, menor será o desvio padrão das estimativas dos efeitos.

5. REFERÊNCIAS BIBLIOGRÁFICAS

ALCOCHETE, A. A. N. de; RANGEL, P. H. N.; FERREIRA, M. E. Mapping of quantitative trait loci for thermosensitive genic male sterility in *indica* rice. **Pesquisa agropecuária brasileira**, v.40, n.12, p.179-88, 2005.

BARBOSA, L.; LOPES, P. S.; REGAZZI, A. J.; GUIMARÃES, S. E. F.; TORRES, R. A. Seleção de variáveis de desempenho de suínos por meio da análise de componentes principais. **Arq. Bras. Med. Vet. Zootec.**, v.57, n.6, p.805-810, 2005.

BARROS, W. S. **Genotipagem seletiva e outras estratégias de genotipagem no mapeamento genético e na detecção de QTLs em populações F₂ simuladas**. Viçosa, MG: UFV, 2007. 158p. (Tese de Doutorado em Genética e Melhoramento). Universidade Federal de Viçosa, Viçosa.

BHERING, L. L. **Mapeamento genético em famílias simuladas de irmãos completos**. Viçosa, MG: UFV, 2008. 150p. (Tese de Doutorado em Genética e Melhoramento). Universidade Federal de Viçosa, Viçosa.

BHERING, L. L.; CRUZ, C. D. Tamanho de população ideal para mapeamento genético em famílias de irmãos completos. **Pesquisa agropecuária brasileira**, v.43, p.379-385, 2008.

BOONCHANAWIWAT, A.; BOONSENG, S. S. O.; LIGHTFOOT, D. A.; TRIWITAYAKORN, K. QTL underlying plant and first branch height in cassava (*Manihot esculenta* Crantz). **Field Crops Research**, v.121, p.343-349, 2011. DOI: 10.1016/j.fcr.2010.12.022

BOROVSKY, Y.; PARAN, I. Characterization of *fs* 10.1, a major QTL controlling fruit in elongation in *Capsicum*. **Theor Appl Genet**, 2011. doi: 10.1007/s00122-011-1615-7

Carlborg, Ö.; Koning, D. J. de; Manly, K. F.; Chesler, E.; Williams, R. W.; Haley, C. S. Methodological aspects of the genetic dissection of gene expression. **Bioinformatics**, v., n.10, 2383-2393, 2005. DOI: 10.1093/bioinformatics/bti241

CHAKRABORTY, S.; ZENG, Z. B. QTL mapping for Days to Flowering under Drought Condition in Rice (*Oryza sativa* L.) Genome. **Not Bot Hort Agrobot Cluj**, v.39, n.1, p.58-63, 2011.

COUTO, K. R.; SANTOS, J. B.; RAMALHO, M. A. P.; SILVA, G. S. Identificação de marcadores microssatélites relacionados ao escurecimento do grão de feijão. **Pesquisa Agropecuária Brasileira**, v.45, n.11, p.1268-1274, 2010.

CRUZ, C. D. **Programa para análise de dados moleculares e quantitativos – GQMOL - Versão 2008.6.1**. Viçosa: UFV, 2008.

Cruz, C.D. **Programa Genes: Estatística experimental e matrizes**. Editora UFV. Viçosa, MG. 285p. 2006

CRUZ, E. M. **Efeito da saturação e do tamanho da populações F2 e de retrocruzamento sobre a acurácia do mapeamento genético.** Viçosa, MG: UFV, 2006. (Tese de Doutorado em Genética e Melhoramento). Universidade Federal de Viçosa, Viçosa.

DING, J. Q.; MA, J. L.; ZHANG, C. R.; DONG, H. F.; XI, Z. Y.; XIA, Z. L.; WU, J. Y. QTL mapping for test weight by using F_{2:3} population in maize. **Journal of Genetics**, v.90, n.1, p.75-80, 2011.

ETGES, W.J.; OLIVEIRA, C.C.; RITCHIE, M.G.; NOO, M.A.F. Genetics of incipient speciation in *Drosophila mojavensis*: ii. host plants and mating status influence cuticular hydrocarbon qtl expression and g × e interactions. **Evolution**, v.63, p.1712-1730, 2009.

FARIA, L. M.. **Mapeamento genético e detecção de QTLs em um cruzamento de limão cravo e citrumelo swingle.** Campinas, SP: IAC, 2007. 93p. (Dissertação de mestrado em Agricultura Tropical e Subtropical, área de concentração em Genética e Melhoramento Genético Vegetal). Instituto Agrônômico, Campinas.

Flint-Garcia, S. A.; Thuillet³, A. C.; Yu, J.; Pressoir, G.; Romero, S. M.; Mitchell, S. E.; Doebley, J.; Kresovich, D.; Goodman, M. M.; Buckler, E. S. Maize association population: a high-resolution platform for quantitative trait locus dissection. **The Plant Journal**, v.44, p.054-1064, 2005. DOI: 10.1111/.1365-313X.2005.02591.x

GONÇALVES, T. de M.; COSTA, A. L. L. da; LARANJO, J. S.; RODRIGUES, M. A P.; REBOUÇAS, G. F. Detecção de locos de características quantitativas (QTLs) afetando o crescimento e a carcaça de suíno: um enfoque Bayseano com o uso de diferentes prioris. **Revista Brasileira de Zootecnia**, v.35, n.2, p.261-272, 2008.

LANDERS, S.; BOTSTEIN, D. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. **Genetics**, v.121, n.1, p.185-199, 1989.

LAPERUTA, L. D. C. **Estudo de uma população segregante (F₁) de maracujá doce: enriquecimento do mapa de ligação e mapeamento de QTL para produção e qualidade de frutos.** Piracicaba, SP: ESALQ, 2011. 134p. (Tese de doutorado em Ciências). Universidade de São Paulo, Piracicaba.

LIU, R.; WANG, B.; GUO, W.; WANG, L.; ZHANG, T. Differential gene expression and associated QTL mapping for cotton yield based on a cDNA-AFLP transcriptome map on immortalized F₂. **Theor. Appl. Genet.**, 2011. DOI: 10.1007/s00122-011-1597-5

LU, B.; XIE, K.; YANG, C.; ZHANG, L.; WU, T.; LI, L.; LIU, X.; JIANG, L.; WAN, J. Efficient QTL detection for heading date in bechcross inbred line and F₂ population derived from the same rice cross. **African Journal of Agriculture Research**, v.6, n.10, p.2372-2378, 2011.

LYNCH, M.; WALSH, B. **Genetics and Analysis of Quantitative Traits.** Sunderland, MA: Sinauer Associates, 1998. 980 p.

MCCLOSKEY, B.; MA, X.; TANKSLEY, S.D. Quantifying the Relative Contribution of the Heterozygous Class to QTL Detect Power. **Statistical Application in Genetics and Molecular Biology**, v.10, n.1, 2011. doi: 10.2202/1544-6116.1622

MIGNON, G.L.; PITEL, F.; GILBERT, H.; BIHAN-DURVAL, E.L.; VIGNOLES, F.; DEMEURE, O.; LAGARRIGUE, S.; SIMON, J.; COGBURN, L.A.; AGGREY, S.E.; DOUAIRE, M.; ROY, P.L. A comprehensive analysis of QTL for abdominal fat and breast muscle weights on chicken chromosome 5 using a multivariate approach. **Animal Genetics**, v.40, p.157-164, 2009.

MORETZSOHN, M. C. **Desenvolvimento e mapeamento de marcadores microssatélites e identificação de QTLs ligados à produtividade e à resistência à mancha preta em *Arachis spp.*** Brasília, DF: UnB, 2011. 116p. (Tese de Doutorado em Biologia Celular). Universidade de Brasília, Brasília.

MORO, G. V. **Uso de seleção genômica e fenotípica de linhagens para a predição de testecrosses em milho.** Piracicabe, SP: ESALQ, 2011. 116p. (Tese de Doutorado em Ciências). Universidade de São Paulo, Piracicaba.

PALHARES, A. C. **Mapeamento genético de marcadores AFLP e de retrotransposons em cana-de-açúcar (*Saccharum spp.*)**. Piracicabe, SP: ESALQ, 2010. 97p. (Tese de Dissertação em Ciências). Universidade de São Paulo, Piracicaba.

PIRES, A. V.; LOPES, P. S.; GUIMARÃES, S. E. F.; PEIXOTO, J. O. Mapeamento de locos de característica quantitativa associados à composição de carcaça, no cromossomo de seis suínos. **Arq. Bras. Med. Vet. Zootec.**, v.60, n.3, p.725-732, 2008.

PIRES, A.V.; LOPES, P.S.; GUIMARÃES, S.E.F.; GUIMARÃES, C.T.; GOMIDE, L.A.M.; BENEVENUTO JÚNIOR, A.A.; CARMO, F.M.S. Quantitative trait loci mapping for meat quality traits in swine chromosome 6. **Arq. Bras. Med. Vet. Zootec.**, v.57, p.608-615, 2005.

RODRIGUES, J. I. S.; MIRANDA, F. D.; FERREIRA, A.; BORGES, L. L.; FERREIRA, M. F. S.; GOOD-GOD, P. I. V.; PIOVESAN, N. D.; BARROS, E. G.; CRUZ, C. D.; MOREIRA, M. A. Mapeamento de QTL para conteúdo de proteína e óleo em soja. **Pesquisa Agropecuária Brasileira**, v.45, n.5, p.472-480, 2010.

SANTOS, R. M. F.; LOPES, U. V.; BAHIA, R. C.; MACHADO, R. C. R.; AHNERT, D.; CORREA, R. X. Marcadores microssatélites relacionados com a resistência à vassoura-de-bruxa do cacauzeiro. **Pesquisa agropecuária brasileiro**, v.42, n.8, p.1137-1142, 2007.

SCHUSTER, I.; CRUZ, C.D. **Estatística genômica aplicada a populações derivadas de cruzamento controlados.** 2º Ed. Viçosa, 2008. 568p.

SILVA, H. D.; VENCOSKY, R. Poder de detecção de “Quantitative Trait Loci”, da análise de marca simples e da regressão linear múltipla. **Scientia Agricola**, v.59, n.4, p.755-762, 2002.

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

SILVA, F. E. J.; GUIDO, L. N.; GAZAFFI, R.; GARCIA, A. A. F.; LEDUR, M. C.; COUTINHO, L. L.; ROSÁRIO, M. F. Regiões genômicas associadas a características de desempenho e carcaça no cromossomo 5 de linhagens brasileiras de galinhas. **Pesquisa Agropecuária Brasileira**, v.46, n.3, p.229-238, 2011.

SOUSA, K. R. S.; GUIMARÃES, S. E. F.; SILVA FILHO, M. I.; LOPES, M. S.; PINTO, A. P. G.; VERANDO, L. L.; BRACCINI NETO, J.; LOPES, P. S. Mapeamento de locos de características quantitativas nos cromossomos 5, 7 e 8 de suínos. **Revista Brasileira de Zootecnia**, v.40, n.1, p.115-123, 2011.

TOLEDO, E.R.; LEANDRO, R.A.; SOUZA JUNIOR, C.L.; SOUZA, A.P. Mapeamento de qtls : uma abordagem bayesiana. **Rev. Bras. Biom.**, v.26, p.107-114, 2008.

TOMÉ, L. G. O. **Análise biométrica e mapeamento de QTLs para tolerância à seca em milho**. Viçosa, MG: UFV, 2005. 148p. (Tese de Doutorado em Biologia Celular). Universidade Federal de Viçosa, Viçosa.

TORTEREAU, F.; GILBERT, H.; HEUVEN, H. C. M.; BIDANEL, J. P.; GROENEM, M. A. M.; RIQUET, J. Number and mode of inheritance of QTL influencing backfat thickness on SSC2p in sino-european pig pedigrees. **Genetics Selection Evolution**, v.43, n.11, 2011. doi: 10.1186/1297-686-43-11

XU, M. H.; WEI, C. S.; TANG, X. T.; ZHU, Z. H.; SIMA, Y. F.; LOU, X. Y. A new mapping method for quantitative trait loci of silkworm. **Genetics**, 2011. DOI: 10.1186/1471-2156-12-19

WU, R.; MA, C. X.; CASALLA, G. **Statistical Genetics of Quantitative Traits: Linkage, Maps, and QTL**. New York, NY: Springer, 2007. 365 p.

YUAN, J.; WANG, B.; WU, L. Quantitative trait loci (QTL) mapping for inflorescence length traits in *Lablab purpureus* (L.) sweet. **African Journal of Biotechnology**, v.10, n.18, p.3558-3566, 2011.

YAN, A.; ZHI-QIU, H.; ZAI-XIANG, T.; XUE-FENG, W.; CHEN-WU, X. A General Method for QTL Mapping in Multiple Related populations Derived from Multiple Parents. **Rice Science**, v. 16, n. 1, 2009.

YOUNG, N. D. **Constructing a plant genetic linkage map with DNA markers**. In: PHILIPS, P. L.; VASIL, I. K. DNA-Based Markers in plants. Dordrecht, The Netherlands: Kluwer Academic Publisher, p.39-57, 1994.

CAPÍTULO 2

EFICIÊNCIA NA DETECÇÃO DE VÁRIOS QTLs EM UM GRUPO DE LIGAÇÃO GÊNICA

RESUMO

FERREIRA, Eliel Alves, D.Sc., Universidade Federal de Viçosa, fevereiro de 2012. **Eficiência na detecção de vários QTLs em um Grupo de Ligação gênica.** Orientador: Cosme Damião Cruz. Coorientadores: Pedro Crescêncio Souza Carneiro e Leonardo Lopes Bhering.

Para o melhoramento genético, tem-se buscado, cada vez mais, a manipulação assistida por marcadores moleculares e esses marcadores também são importantes nos estudos genéticos. Dessa forma, são ferramentas úteis para detectar variações no genoma, aumentando o poder das análises genômicas. Um problema comum nos estudos de QTL é a detecção de QTL fantasma, quando há mais de um QTL no mesmo ou em outros grupos de ligação. Assim, o objetivo do presente trabalho foi avaliar as metodologias de marca simples, intervalo simples e composto em relação à detecção e posicionamento de QTL, além da detecção de QTLs fantasmas. Os dados foram gerados utilizando o módulo de simulação do aplicativo computacional GQMOL. Foram simulados genomas parentais e amostras de populações F_2 , contendo 1000 indivíduos cada e foram simuladas 50 repetições com as mesmas características. O genoma dessas populações foi simulado tomando como referência uma espécie fictícia diplóide com $2n = 2x = 8$ cromossomos, composto de quatro grupos de ligação. Cada grupo de ligação formado possuiu tamanho de 100 cM, contendo 21 marcadores codominantes com saturação de 5 cM. Estipulou-se que os grupos de ligação 1, 2 e 3 continham dois QTLs, explicando (28% e 28%), (28% e 5%) e (5% e 5%) da variação genotípica, respectivamente, e que o grupo de ligação 4 não conteria QTL, sendo utilizado como controle. Para cada repetição foi simulada uma característica com média igual a 100, herdabilidade (h^2) igual a 50% e coeficiente variação ambiental (Cv) igual a 10%. O grau médio de dominância foi considerado igual a 1. Posteriormente, foram realizados os testes de segregação e as análises de detecção de QTL. Para estes foram utilizadas as metodologias de marca simples, intervalo simples e composto, quando necessário. Observou-se que ao aplicar a metodologia de intervalo composto melhorou-se a detecção de QTL. Dessa forma, conclui-se que a metodologia de intervalo composto melhorou a não detecção de QTL fantasma e que o QTL que explica mais a característica suprime o que explica menos, quando comparado à metodologia de intervalo simples.

Palavras chaves: População de cruzamento controlado; QTL inexistente; populações simuladas.

ABSTRACT

FERREIRA, Eliel Alves, D.Sc., Universidade Federal de Viçosa, february de 2012. The **Efficient in the many QTLs detection in the gene linkage group**. Adviser: Cosme Damião Cruz. Co-advisers: Pedro Crescêncio Souza Carneiro and Leonardo Lopes Bhering.

Genetic improvement has been sought the manipulation assisted by molecular markers and these markers are also important in genetic studies. Thus, they are useful tools to detect genomic variations, increasing the power of genomic analysis. A common problem in QTL studies is the ghost QTL detection when more than one in the same or other linkage groups. The objective of this study was to evaluate the methodologies of the simple mark, simple and composite interval to detection and positioning of QTL, and ghost QTL detection. The parental genomes and F₂ populations were simulated by the GQMOL software. The population contained 1000 individuals and 50 replicas were simulated. The genome was simulated for diploid species with $2n = 2x = 8$ chromosomes, composed of five linkage groups. Each linkage group was of 100 cM, containing 21 co-dominant markers with saturation of 5 cM. Stipulated that the linkage groups 1, 2 and 3 contained two QTL, explaining (28% and 28%), (28% and 5%) and (5% and 5%) of the phenotypic variation, respectively, and the linkage group 4 don't contained QTL. Each replica was simulated with yield of 100, heritability (h^2) of 50% and coefficient environmental variation (CVe) of 10%. The average degree of dominance was considered equal to 1. Later, were applied segregation tests and QTL analysis. For QTL analysis was used the simple marking methodologies, simple and composite interval. It was observed that with applying the composite interval method has improved QTL detection, decreasing the ghost QTL detection. It was found that with the interval composite method avoided that the QTL major effect supreme the effect of QTL small. Therefore, it can be concluded that the composite interval methodology is efficient for QTL detection and prevention of ghost QTL.

Key words: Population-controlled crossover; QTL nonexistent; simulated populations.

1. INTRODUÇÃO

No melhoramento genético, tem-se buscado, cada vez mais, a manipulação assistida por marcadores moleculares, visando obter maior eficiência nas transferências genéticas, tanto em relação às características qualitativas quanto quantitativas. Marcadores moleculares ligados a características de importância econômica têm sido identificados, permitindo a seleção indireta em gerações cada vez mais precoces. Esta estratégia reduz tempo e recursos necessários, não só para desenvolver populações segregantes por várias gerações, mas também para estimar parâmetros.

Além de sua importância no melhoramento, os marcadores também são importantes nos estudos genéticos. Neste contexto, os marcadores moleculares são ferramentas úteis para detectar variações no genoma, aumentando o poder da análise genética. Atualmente, existe grande variedade de marcadores moleculares disponíveis para diferentes espécies vegetais, sendo que os principais são: RFLP (*Restriction fragment Length Polymorphism*), RAPD (*Random Amplified Polymorphic DNA*), SSR (*Simple Sequence Repeats*), AFLP (*Amplified Fragment Length Polymorphism*), SNPs (*Single Nucleotide Polymorphisms*) e os DArT (*Diversity Arrays Technology*) (CAIXETA et al., 2009; OLIVER et al., 2011).

Os marcadores são classificados como dominantes ou codominantes, sendo que ambos podem ser utilizados para construir mapas de ligação. Esses mapas auxiliam o melhoramento genético, principalmente na detecção e posicionamento de locos quantitativos. Esses QTLs podem ser detectados por metodologias que utilizam contrastes entre médias (ANOVA e teste t). Os marcadores podem ser associados ao QTL por meio do método da marca simples baseada na regressão linear ou na verossimilhança, no entanto, estes métodos não posicionam o QTL no grupo de ligação, apesar de ser eficiente em sua detecção. Quando o interesse for, além de sua detecção, localizar o seu posicionamento, é possível utilizar os métodos de intervalo, podendo ser o método de intervalo simples e/ou intervalo composto, sendo que este último utiliza marcadores como cofatores. Essas metodologias podem ser baseadas em regressão linear ou em verossimilhança.

A metodologia de intervalo simples detecta e posiciona o QTL de forma eficiente, principalmente quando se utiliza populações grandes. Entretanto, este método avalia um intervalo de cada vez sem considerar o restante da informação do genoma, podendo apresentar problemas na detecção quando existem outros QTLs no mesmo ou

em outros grupos de ligação. Dessa forma, não há como determinar se o efeito significativo observado é devido a um ou a vários QTLs. Além disso, quando múltiplos QTLs estão presentes, um problema comum é a observação de falsos positivos, denominados de QTLs fantasmas (SCHUSTER; CRUZ, 2008; JANGARELLI et al., 2011).

Do exposto, o objetivo do presente trabalho foi avaliar as metodologias da marca simples, intervalo simples e composto em relação à detecção e posicionamento do QTL, além da detecção de QTLs fantasmas.

6. MATERIAL E MÉTODOS

2.1. Simulação

2.1.1. Simulação dos dados

Para gerar os dados foi utilizado o módulo de simulação do aplicativo computacional GQMOL (CRUZ, 2008), que permite obter informações sobre os genomas, genótipos, genitores, indivíduos de diferentes tipos de populações e dados de características quantitativas. Foram simulados genomas parentais e amostra de populações de cruzamento controlado (F_2) com tamanho de 1000 indivíduos.

Foram simuladas 50 repetições com as mesmas características. O genoma dessas populações foi simulado tomando como referência uma espécie fictícia diplóide com $2n = 2x = 8$ cromossomos, composto de quatro grupos de ligação (GL). Cada grupo de ligação formado possuiu tamanho de 100 cM, contendo 21 marcadores codominantes com saturação de 5 cM.

2.1.2. Simulação dos genitores

Para a simulação dos genitores foram considerados os pais homozigotos contrastantes e os mapas genéticos utilizados para obtenção das populações estão representados na Figura 9.

2.1.3. Procedimentos para simulação dos indivíduos F_2

A estratégia básica de simulação é caminhar ao longo dos cromossomos, realizando permutas em cada intervalo entre marcas adjacentes, de acordo com as distâncias dos marcadores, conforme descrito por Silva (2005).

O processo de simulação das populações F_2 seguiu os seguintes passos:

3. A partir do genoma simulado foram construídos os genótipos parentais, de forma que se apresentem os genitores homozigotos contrastantes sempre de padrão genotípico $A_iA_i \times a_ia_i$.
4. A partir dos genótipos parentais foram gerados os gametas para a formação dos indivíduos da população F_2 . As produções dos gametas foram feitas simulando-se o pareamento dos homólogos e realizando-se permutas ao longo do cromossomo; considerou-se a não interferência nas regiões delimitadas por dois marcadores adjacentes. A probabilidade de ocorrência de recombinação numa região entre marcadores adjacentes é dada de acordo com a distância destes marcadores no genoma simulado. Sendo que uma maior distância implica numa maior possibilidade de ocorrência de recombinação.

O programa GQMOL considera o encontro aleatório de gametas para a simulação dos indivíduos. Sendo assim, um novo processo acontece para cada indivíduo simulado dentro de cada repetição.

2.1.5. Simulação da característica quantitativa (QTL)

Foi estipulado que os grupos de ligação 1, 2 e 3 conteriam dois QTLs. Também foi estabelecido que o grupo de ligação 4 não conteria QTL, sendo utilizado como controle da metodologia aplicada à detecção de QTL. Para cada repetição foi simulada a característica com média igual a 100, herdabilidade (h^2) igual a 50% e coeficiente variação ambiental (CVe) igual a 10%.

Considerou-se também que a contribuição gênica dos QTLs foram diferentes em cada GL, sendo os QTLs posicionados nos GL 1, 2 e 3 explicariam: (28% e 28%), (28% e 5%) e (5% e 5%) da variação genotípica, respectivamente. O grau médio de dominância foi considerado igual a 1, ou seja, dominância completa. Os mapas genéticos, com o posicionamento dos QTLs simulados estão apresentadas na Figura 10.

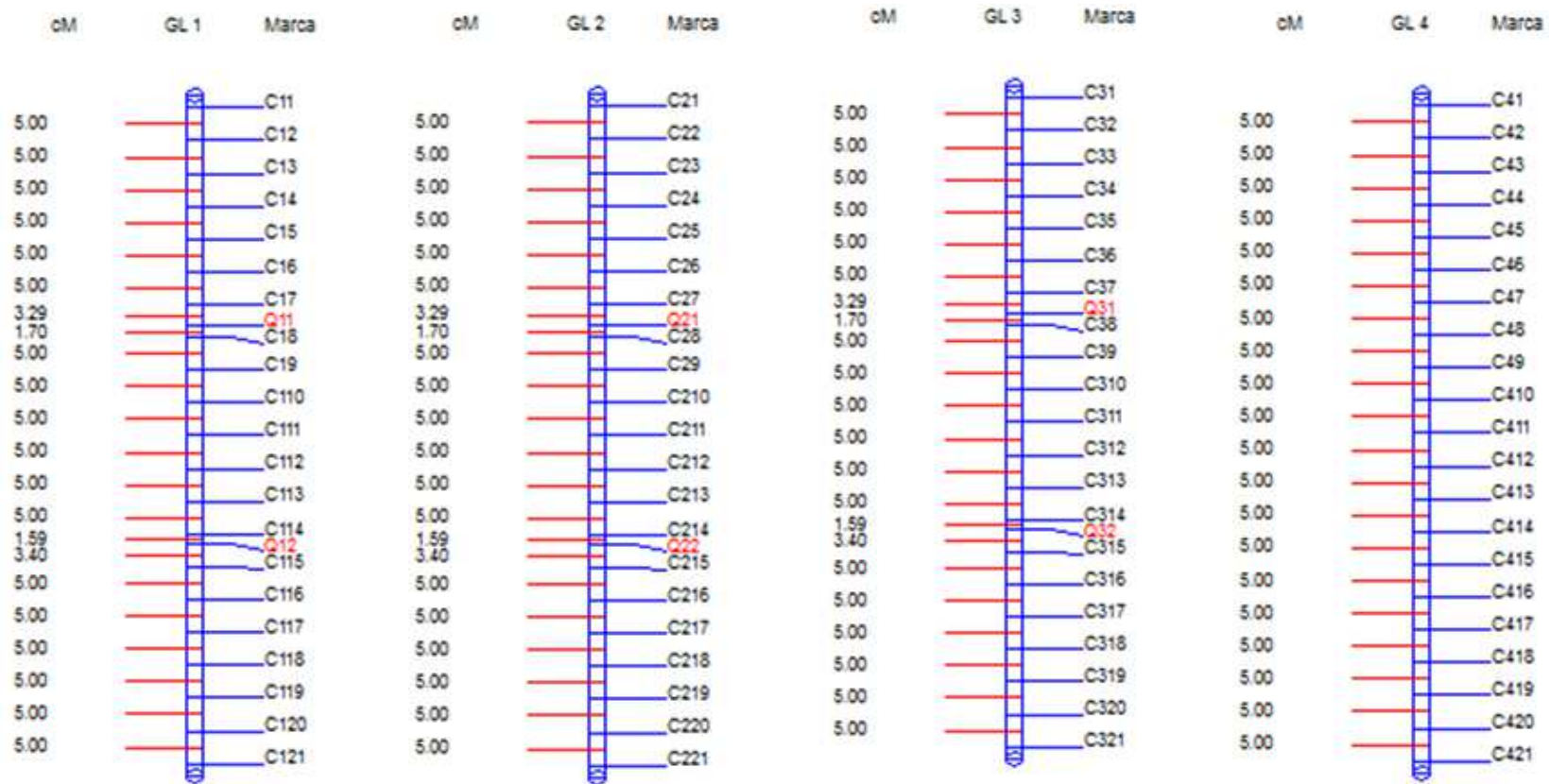


Figura 10. Mapa genético dos quatro grupos de ligação da população F_2 simulada com o posicionamento dos QTLs estabelecidos.

2.2. Mapeamento

Após as populações serem geradas, seguiram-se as etapas do processo de mapeamento, como descrito a seguir.

2.2.1. Análise de segregação

Foram aplicados testes de qui-quadrado (χ^2) para verificar se há segregação em cada marca, sendo estas esperadas de 1:2:1.

A estatística utilizada para o teste de segregação foi a qui-quadrado, dada por:

$$\chi^2 = \sum_{i=1}^n \left[\frac{(Obs_i - Esp_i)^2}{Esp_i} \right]$$

sendo,

χ^2 é o valor de qui-quadrado calculado;

Obs_i e Esp_i são valores observados e esperados, para i-ésima classe fenotípica (i = 1, 2 e 3).

A hipótese (H₀) de segregação dos locos foi testada ao nível conjunto de 5% de probabilidade e dois graus de liberdade. Nas situações em que o valor da probabilidade calculada foi inferior ao pré-estabelecido, a hipótese H₀ foi rejeitada significando que a segregação não ocorreu de acordo com o esperado.

Como foi testado um grande número de marcadores, utilizou-se a proteção de Bonferroni para reduzir a probabilidade de se cometer o erro tipo I, mantendo a significância total α_T igual a 5%. Com o uso da correção de Bonferroni o nível de significância para teste individual é dada por:

$$\alpha = -\exp\left(\frac{\ln(1 - \alpha_T)}{N}\right) + 1$$

A utilização da proteção de Bonferroni tem por objetivo fixar um nível crítico de significância α_T para o experimento. Para isso, deve-se estimar um nível de significância α para cada teste que proporcione o nível de significância α_T desejado para cada experimento (CRUZ; SILVA, 2009). Esta análise foi realizada por meio do software QMOL (CRUZ, 2008).

2.2.2. Estimação da porcentagem de recombinação

Após a aplicação dos testes de segregação, seguiu-se a etapa da estimação da porcentagem de recombinação entre os pares de marcas, utilizando-se o método da máxima verossimilhança. Os estimadores de verossimilhança foram obtidos pelo método gráfico em que valores da razão de verossimilhança (LR) ou LOD são analisados graficamente, representativos do eixo Y, associados aos valores da porcentagem de recombinação, no eixo X. A função de verossimilhança é dada por:

$$L(r; n_{ij}) = \lambda p_1^{n_{11}} p_2^{n_{12}} p_3^{n_{13}} p_4^{n_{21}} p_5^{n_{22}} p_6^{n_{23}} p_7^{n_{31}} p_8^{n_{32}} p_9^{n_{33}}$$

sendo,

r é a porcentagem de recombinação;

$$\lambda = \frac{N!}{n_{11}! n_{12}! \dots n_{33}!};$$

$$N = \sum_{i=1}^3 \sum_{j=1}^3 n_{ij};$$

P_i é a probabilidade associada a cada classe;

n_{ij} é o número de indivíduos observados em cada probabilidade.

As probabilidades associadas a cada classe, cujas fases de ligação podem ser de aproximação ou repulsão, juntamente com o número de indivíduos observados, estão apresentadas na Tabela 41.

Tabela 41. Frequências genótípicas esperadas na população F₂, utilizando marcadores codominantes no cruzamento entre indivíduos com genitores homocigotos contrastantes.

Genótipo	Fase de aproximação			Fase de repulsão		
	Código*	Núm. obs.	Frequência esperada	Código*	Núm. obs.	Frequência esperada
AABB	0-0	n ₁₁	p ₁ =1/4(1-r) ²	0-2	n ₁₃	p ₁ =1/4r ²
AABb	0-1	n ₁₂	p ₂ =1/2r(1-r)	0-1	n ₁₂	p ₂ =1/2r(1-r)
AAbb	0-2	n ₁₃	p ₃ =1/4r ²	0-0	n ₁₁	p ₃ =1/4(1-r) ²
AaBB	1-0	n ₂₁	p ₄ =1/2r(1-r)	1-2	n ₂₃	p ₄ =1/2r(1-r)
AaBb	1-1	n ₂₂	p ₅ =1/2(1-r)+1/2r ²	1-1	n ₂₂	p ₅ =1/2(1-r)+1/2r ²
Aabb	1-2	n ₂₃	p ₆ =1/2r(1-r)	1-0	n ₂₁	p ₆ =1/2r(1-r)
aaBB	2-0	n ₃₁	p ₇ =1/4r ²	2-2	n ₃₃	p ₇ =1/4(1-r) ²
aaBb	2-1	n ₃₂	p ₈ =1/2r(1-r)	2-1	n ₃₂	p ₈ =1/2r(1-r)
aabb	2-2	n ₃₃	p ₉ =1/4(1-r) ²	2-0	n ₃₁	p ₉ =1/4r ²

*Codificação numérica de cada indivíduo, para os dois locos.

O LOD score é um teste de razão de verossimilhança que utiliza o logaritmo na base 10 (Log₁₀) da razão entre a verossimilhança considerando o valor de r estimado e a verossimilhança para r igual a 0,5.

$$LOD = \text{Log}_{10} \left(\frac{L(r;n_{ij})}{L(r=0,5;n_{ij})} \right)$$

Adotou-se, no presente trabalho, o LOD mínimo de 3, indicando que a verossimilhança, considerando o valor de r, é 1000 vezes maior que a verossimilhança em caso de ligação.

2.2.4. Determinação dos grupos de ligação

Na formação de grupo de ligação utilizou-se a propriedade transitiva, ou seja, se o loco A está ligado ao loco B, e o loco B está ligado ao loco C, logo o loco A está ligado ao loco C, independentemente da frequência de recombinação A e C, portanto,

A, B e C pertencem ao mesmo grupo de ligação. Os critérios utilizados no agrupamento foram a frequência máxima de recombinação (r_{\max}) e o LOD mínimo (LOD_{\min}), para inferir se dois locos estão ligados. Foram utilizados os valores 30% e 3, respectivamente, para r_{\max} e LOD_{\min} . As marcas mais próximas, em relação às marcas já incorporadas, que atenderem aos dois critérios adotados serão incorporadas ao grupo de ligação. Assim, o processo continua, investigando quanto à porcentagem de recombinação e ao LOD entre marcas é entre possíveis vizinhos a serem incorporados às extremidades do grupo de ligação em formação.

2.2.5. Ordenamento das marcas no grupo de ligação

Após a formação dos grupos de ligação foi detalhada a melhor ordem, das n marcas nos grupos.

Um dos processos é gerar todas as possíveis ordens ($n!/2$) e adotar como critério a identificação da melhor ordem, sendo aquela que proporcionar menor soma entre distâncias. Contudo, na medida em que o número de locos marcadores aumenta, aumenta-se grandemente o número de ordens possíveis, consumindo, portanto, um grande tempo de processamento. Então, foi utilizado, alternativamente, o critério da soma das frações de recombinação adjacente (SARF – Sum of Adjacent Recombination Fraction) para os mapas construídos sem análise multiloco. Neste processo, a melhor ordem é aquela que apresenta menor soma de recombinações adjacentes. Considera-se a ordem original estabelecida pelo processo de agrupamento e aplica-se o algoritmo RCD (Rapid Chain Delineation), que consiste em realizar permutas entre dois marcadores vizinhos ou distantes envolvendo três ou quatro marcadores. A ordem é alterada se, após a permuta, a soma das distâncias adjacentes for reduzida. Após todas as permutas conclui-se que a melhor ordem é aquela de menor soma de distâncias adjacentes (SCHUSTER; CRUZ, 2008).

2.3. Análise de detecção de QTLs

Os métodos utilizados neste trabalho, para detecção de QTL, foram os métodos da marca simples, intervalo simples e intervalo composto. Todas essas análises foram realizadas utilizando o aplicativo computacional GQMOL (CRUZ, 2008) que pode ser encontrado no endereço <http://www.ufv.br/dbg/biodata.html>.

2.3.1.1. Método da Marca simples

O método da marca simples foi empregado para avaliar a associação entre genótipos e marcadores moleculares, utilizando-se regressão linear. Para a análise de regressão, os três genótipos (AA, Aa e aa) foram codificados como 1, 0 e -1 para quantificação do efeito aditivo e como 0, 1 e 0 para quantificação do efeito atribuído à dominância, utilizando o modelo:

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

sendo,

Y_j é o valor da característica quantitativa avaliada no j-ésimo indivíduo da população F_2 ;

X_{1j} é o código do marcador para o efeito aditivo (AA = 1, Aa = 0 e aa = -1);

X_{2j} é o código do marcador para o efeito dominante (AA = 0, Aa = 1 e aa = 0);

β_0 é o intercepto (média da característica);

β_1 é o coeficiente associado ao efeito aditivo;

β_2 é o coeficiente associado ao efeito dominante; e

ε_j é o erro aleatório manifestado na característica no j-ésimo indivíduo.

2.3.1.2. Método de intervalo simples

O método de detecção de QTL por intervalo foi proposto por Lander; Botstein (1989) com a finalidade de testar a possibilidade de ocorrência do loco controlador da característica (QTL) de uma série de posições dentro do intervalo flanqueadas por dois marcadores, utilizando a regressão linear para estimação dos parâmetros. Este modelo de regressão, para população F_2 , é dado por:

$$Y_j = \mu + ax_j^* + dz_j^* + \varepsilon_j$$

sendo,

Y_j é o valor da característica Y no indivíduo j;

μ é a média da característica da população;

a é o efeito aditivo do loco que está sendo estudado sobre a característica;

d é o efeito de dominância do loco que está sendo estudado sobre a característica;

x_j^* e z_j^* são as variáveis condicionadoras e dependentes dos genótipos dos marcadores que flanqueiam o QTL, no indivíduo j; e

ε_j é o erro aleatório.

2.3.1.3. Método de intervalo composto

No mapeamento de QTL, os testes que avaliam um intervalo de cada vez, sem considerar o restante da informação do genoma, podem apresentar problemas de detecção, principalmente se houver outros QTLs de grande efeito no mesmo grupo ou em outros grupos de ligação. Isso ocorre com o mapeamento por intervalo simples, cuja análise baseia-se exclusivamente nos marcadores que flanqueiam o intervalo avaliado (SCHUSTER; CRUZ, 2008).

A inclusão de marcadores que não estão associados com a característica pode levar a um aumento na variância residual, ao invés de redução. Portanto, o ideal é que se utilize apenas um subconjunto de marcadores selecionados como cofatores (c), que serão utilizados na análise. Os cofatores considerados nas análises foram selecionados utilizando-se o procedimento de regressão de *stepwise*, com $\alpha = 0,05$ para inclusão e exclusão dos marcadores no modelo (SCHUSTER; CRUZ, 2008; CRUZ et al., 2009).

Assim, para um melhor entendimento, utilizou-se o mapeamento de QTLs por intervalo composto (ZENG, 1993). O modelo do método de intervalo composto é dado por:

$$Y_j = [\mu + ax_j^* + dz_j^*] + \sum_{k \neq i, i+1}^c b_k x_{kj} + \varepsilon_j$$

sendo,

i e $i+1$ são os valores dos marcadores no intervalo observado;

y_j é o valor da característica y no indivíduo j ;

x_{kj} é o escore do marcador k no indivíduo j (MM=1, Mm=0, mm=-1);

b_k é o coeficiente de regressão parcial do fenótipo Y com o k -ésimo marcador;

ε_j é o erro aleatório;

O termo entre colchetes é o mesmo apresentado no modelo de regressão para mapeamento de intervalo simples.

2.4. Eficiência na obtenção das estimativas dos efeitos e posicionamento dos QTLs

Foram realizados testes t de Student, entre as médias obtidas pelas análises e as simuladas, dos efeitos das médias genóticas, aditivos, devidos ao desvio de dominância e contribuição gênica (R^2), para o método de marca simples. Esta análise também foi realizada para os efeitos das médias genóticas, aditivos, devidos ao desvio de dominância, contribuição gênica (R^2) e posição do QTL referentes ao método de intervalo simples e composto. Tanto para o método da marca simples quanto para os de intervalo, os teste t foram realizados para os grupos de ligação 1, 2 e 3, pois no GL 4 é sabido que não há presença de QTL. O método do teste t é dado por:

$$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

sendo,

\bar{x} é média dos efeitos;

μ_0 é o valor da constante utilizado para comparação com a média dos efeitos;

s é o desvio padrão amostral;

n é o tamanho da amostra.

Na metodologia da marca simples foi analisada a frequência dos marcadores ligados aos QTLs em cada grupo de ligação. Em relação às metodologias de intervalo foi analisada a distribuição das réplicas em relação à posição do QTL detectado em cada grupo de ligação.

Sabe-se que a localização correta do primeiro QTL simulado é a 33,29 cM e do segundo é 66, 59.

Foi analisada a eficiência na detecção de QTLs fantasmas. Para isto, analisou-se pelo método de intervalo simples e quando a detecção foi diferente de dois QTLs, utilizou-se a metodologia de intervalo composto, visando melhorar a precisão na detecção dos QTLs.

3. RESULTADOS E DISCUSSÃO

3.1. Análise das marcas moleculares

Com o intuito de verificar se as marcas segregam como o esperado (1:2:1), aplicou-se o teste de qui-quadrado (χ^2) com a proteção de Bonferroni, com a significância total (α_T) igual a 5%, com o objetivo de reduzir o erro tipo I (SCHUSTER; CRUZ, 2008).

Na Tabela 42 encontram-se os resultados das análises de segregação para as marcas individuais com a proteção de Bonferroni, referentes à uma repetição. Foram observados que as marcas segregaram na proporção de 1:2:1, como o esperado, sem distorção. O mesmo foi verificado para as demais repetições.

Posteriormente foi realizado o processo de mapeamento das populações, verificando-se que os quatro grupos de ligação foram recuperados, contendo 21 marcadores cada, em todas as repetições. Foi observado também que não houve nenhuma inversão no posicionamento das marcas (Figura 11).

Tais resultados eram esperados, pois o tamanho populacional de 1000 indivíduos recupera os grupos e mantém a ordenação das marcas, como visto no capítulo 1. Resultados semelhantes foram obtidos por Bhering; Cruz (2008) que relataram que populações superiores a 200 indivíduos recuperam e mantêm a ordem dos marcadores moleculares em famílias de irmãos completos, completamente informativa.

Tabela 42. Teste de segregação (Hipótese de segregação igual a 1:2:1) dos 84 marcadores ao nível conjunto de 5% de probabilidade pelo teste χ^2 com proteção de Bonferroni.

Marcador	Classes			Qui- quadrado	Probabilidade(%)
	0	1	2		
C ₁₁	218	514	268	5,78	5,55 ^{ns}
C ₁₂	207	526	267	9,90	0,71 ^{ns}
C ₁₃	216	512	272	6,85	3,26 ^{ns}
C ₁₄	212	527	261	7,72	2,11 ^{ns}
C ₁₅	212	536	252	8,38	1,51 ^{ns}
C ₁₆	225	523	252	3,57	16,75 ^{ns}
C ₁₇	239	512	249	0,78	67,84 ^{ns}
C ₁₈	236	514	250	1,18	55,54 ^{ns}
C ₁₉	234	526	240	2,78	24,96 ^{ns}
C ₁₁₀	239	529	232	3,46	17,71 ^{ns}
C ₁₁₁	231	522	247	2,45	29,40 ^{ns}
C ₁₁₂	235	515	250	1,35	50,92 ^{ns}
C ₁₁₃	242	501	257	0,45	79,69 ^{ns}
C ₁₁₄	252	488	260	0,70	70,33 ^{ns}
C ₁₁₅	241	501	258	0,58	74,75 ^{ns}
C ₁₁₆	245	508	247	0,26	87,63 ^{ns}
C ₁₁₇	257	500	243	0,39	82,20 ^{ns}
C ₁₁₈	259	501	240	0,73	69,56 ^{ns}
C ₁₁₉	268	495	237	2,02	36,38 ^{ns}
C ₁₂₀	265	496	239	1,42	49,26 ^{ns}
C ₁₂₁	264	481	255	1,61	44,80 ^{ns}
C ₂₁	243	527	230	3,25	19,65 ^{ns}
C ₂₂	250	526	224	4,06	13,16 ^{ns}
C ₂₃	254	534	212	8,15	1,70 ^{ns}
C ₂₄	243	538	219	6,93	3,13 ^{ns}
C ₂₅	235	527	238	2,93	23,06 ^{ns}
C ₂₆	234	528	238	3,17	20,51 ^{ns}
C ₂₇	244	519	237	1,54	46,25 ^{ns}
C ₂₈	243	518	239	1,33	51,48 ^{ns}
C ₂₉	248	517	235	1,49	47,38 ^{ns}
C ₂₁₀	243	518	239	1,33	51,48 ^{ns}
C ₂₁₁	249	517	234	1,61	44,80 ^{ns}
C ₂₁₂	257	516	227	2,82	24,37 ^{ns}
C ₂₁₃	253	513	234	1,90	85,00 ^{ns}
C ₂₁₄	253	506	241	0,43	80,57 ^{ns}
C ₂₁₅	243	519	238	1,49	47,38 ^{ns}
C ₂₁₆	241	521	238	1,78	41,02 ^{ns}
C ₂₁₇	233	522	245	2,22	32,89 ^{ns}
C ₂₁₈	238	518	244	1,37	50,46 ^{ns}
C ₂₁₉	251	502	247	0,05	97,63 ^{ns}
C ₂₂₀	249	497	254	0,09	95,79 ^{ns}
C ₂₂₁	259	481	260	1,45	48,53 ^{ns}
C ₃₁	252	494	254	0,15	92,68 ^{ns}

Tabela 42. Continuação.

C ₃₂	256	486	258	0,79	67,30 ^{ns}
C ₃₃	253	477	270	2,69	26,00 ^{ns}
C ₃₄	254	475	271	3,08	21,46 ^{ns}
C ₃₅	267	469	264	3,86	14,50 ^{ns}
C ₃₆	264	484	252	1,31	51,89 ^{ns}
C ₃₇	264	489	247	1,06	58,80 ^{ns}
C ₃₈	255	497	248	0,13	93,52 ^{ns}
C ₃₉	253	506	241	0,43	80,57 ^{ns}
C ₃₁₀	261	489	250	0,73	69,56 ^{ns}
C ₃₁₁	259	499	242	0,58	74,75 ^{ns}
C ₃₁₂	263	499	238	1,25	53,42 ^{ns}
C ₃₁₃	257	514	229	2,35	30,85 ^{ns}
C ₃₁₄	253	515	232	1,78	41,02 ^{ns}
C ₃₁₅	256	507	237	0,92	63,19 ^{ns}
C ₃₁₆	255	499	246	0,17	92,03 ^{ns}
C ₃₁₇	243	522	235	2,06	35,63 ^{ns}
C ₃₁₈	240	529	231	3,53	17,15 ^{ns}
C ₃₁₉	247	514	239	0,91	63,38 ^{ns}
C ₃₂₀	254	498	248	0,09	95,69 ^{ns}
C ₃₂₁	260	497	243	0,61	73,56 ^{ns}
C ₄₁	243	513	244	0,68	71,25 ^{ns}
C ₄₂	248	504	248	0,06	96,85 ^{ns}
C ₄₃	248	511	241	0,58	74,75 ^{ns}
C ₄₄	254	515	231	1,96	37,57 ^{ns}
C ₄₅	245	525	230	2,95	22,88 ^{ns}
C ₄₆	253	513	234	1,40	49,71 ^{ns}
C ₄₇	259	504	237	1,03	59,69 ^{ns}
C ₄₈	269	499	232	2,74	25,38 ^{ns}
C ₄₉	267	500	233	2,31	31,47 ^{ns}
C ₄₁₀	255	506	239	0,66	72,04 ^{ns}
C ₄₁₁	262	493	245	0,77	67,91 ^{ns}
C ₄₁₂	259	490	251	0,53	76,80 ^{ns}
C ₄₁₃	264	495	241	1,16	56,05 ^{ns}
C ₄₁₄	255	511	234	1,37	50,51 ^{ns}
C ₄₁₅	267	505	228	3,14	20,78 ^{ns}
C ₄₁₆	269	517	214	7,21	2,72 ^{ns}
C ₄₁₇	270	511	219	5,69	5,82 ^{ns}
C ₄₁₈	265	516	219	5,26	7,22 ^{ns}
C ₄₁₉	271	513	216	6,73	3,46 ^{ns}
C ₄₂₀	272	503	225	4,45	10,78 ^{ns}
C ₄₂₁	280	501	219	7,45	2,42 ^{ns}

Nível crítico em teste individual igual a 0,0610448 %

^{ns} – Não significativo.

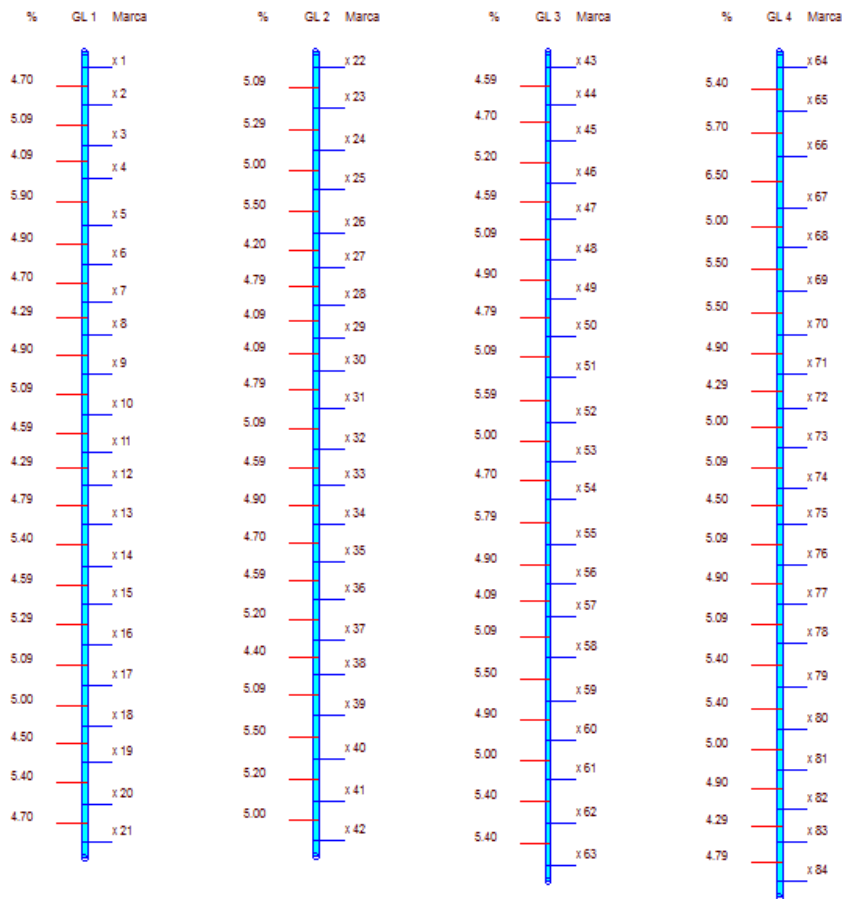


Figura 11. Mapa genético representativo da análise de uma repetição.

3.2. Detecção de QTL

3.2.1. Detecção de QTL por marca simples

É importante salientar que, neste estudo, foi considerado que os grupos de ligação 1, 2 e 3 continham dois QTLs cada, sendo o primeiro na posição 33,29 cM e o segundo na de 66,59 cM. No grupo de ligação 1, os dois QTLs explicavam 28% da variação genotípica, no grupo 2 um explicava 28% e o outro 5% e no grupo 3 encontravam-se dois QTLs explicando 5% da variação genotípica.

Encontram-se na Figura 12 as frequências dos marcadores ligados aos QTLs, sendo que no grupo de ligação 1 foi detectada ligação dos 21 marcadores em todas as repetições. No segundo grupo, três marcadores foram ligados, aos QTLs, em 98% das repetições e os demais marcadores em 100% e no terceiro grupo oito marcadores não foram ligados aos QTLs em todas as repetições. Este fato pode ser devido aos QTLs deste grupo que explicam apenas 5% da característica.

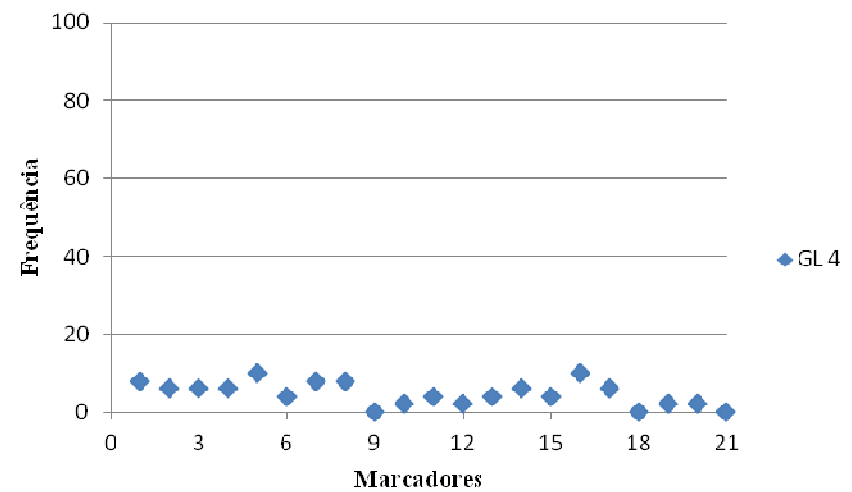
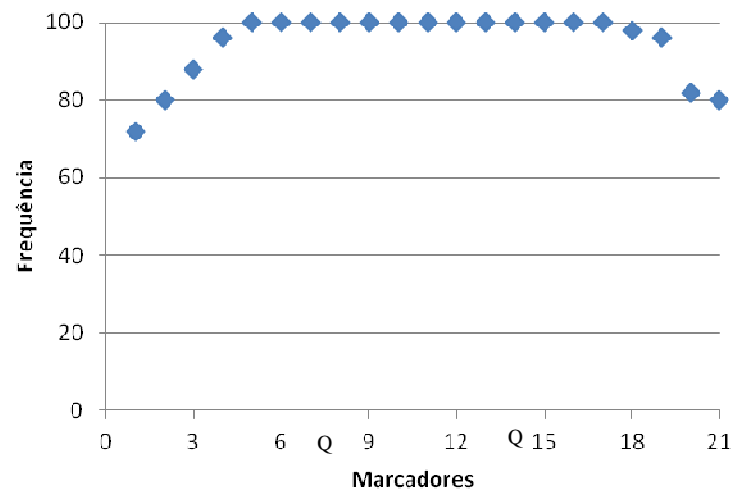
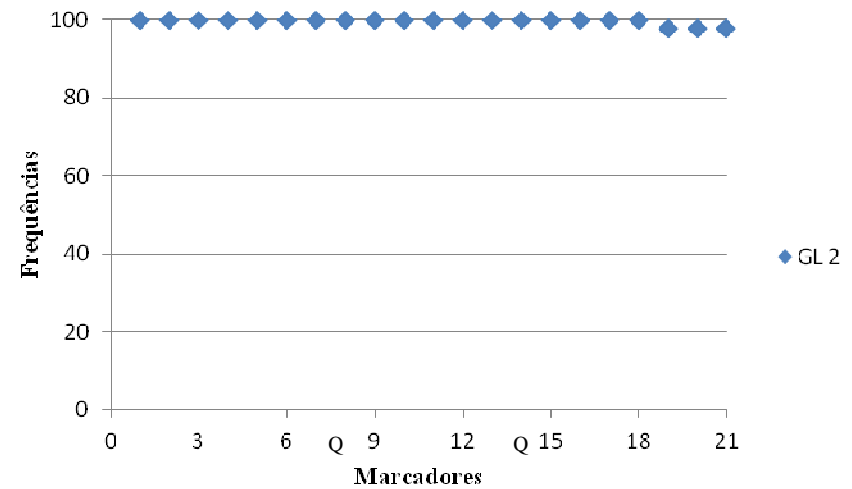
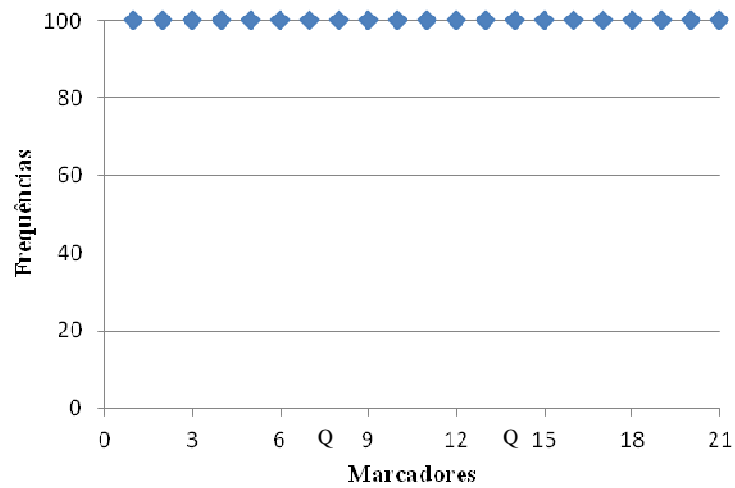


Figura 12. Frequências dos 84 marcadores ligados aos QTLs nos grupos de ligação 1, 2, 3 e 4 obtidas nas populações simuladas de 1000 indivíduos. No eixo das ordenadas estão os valores das frequências em porcentagem e no eixo das abscissas os marcadores moleculares.

Observa-se que os três grupos de ligação contêm dois QTLs e que a maioria dos marcadores está ligada a pelo menos um QTL, portanto, não se sabe qual marcador está ligado ao primeiro ou ao segundo. Quando se analisa o R^2 , de cada marcador, considera-se que o marcador que obtiver o maior R^2 é o que está mais próximo do QTL. Desta forma, foram considerados os marcadores que estão ligados e mais próximos dos QTLs, no mesmo grupo, os dois que possuíram os maiores valores de R^2 .

No Apêndice 1, verifica-se que os marcadores 8, 14, 29, 35, 50 e 56 obtiveram os maiores valores de R^2 com 21,50; 23,07; 21,25; 10,69; 3,44 e 3,11, respectivamente; portanto, são os marcadores mais próximo aos QTLs. Vale ressaltar que os marcadores (8 e 14), (29 e 35) e (50 e 56) são referentes aos grupos de ligação 1, 2 e 3, respectivamente. No Apêndice 2, foi observado que os marcadores (8, 14), 29 e (52 e 57) obtiveram os valores de R^2 de 18,78; 20,57; 29,37; 5,07 e 4,15, respectivamente. Foi verificado que somente o marcador 29 obteve o pico de R^2 , logo, considera-se que contém apenas um QTL e não dois, como foi simulado.

Outro fato a ser observado é que, ao invés de dois picos do R^2 , existem mais, o que mostra conter também mais de dois QTL. Observa-se no Apêndice 3 que os marcadores 9, 12 e 14 obtiveram os picos de R^2 , indicando a presença de três QTLs. Estes marcadores pertencem ao grupo 1 e é sabido que este grupo de ligação continha dois QTLs e não três. A observação do terceiro QTL é um falso positivo, que é denominado de QTL fantasma (LAPERUTA, 2011). A metodologia da marca simples não identifica se esse terceiro QTL é um falso positivo ou um resultado verdadeiro.

Nas Tabelas 43 e 44 estão apresentadas as médias e os valores do teste t referentes aos efeitos das médias genotípicas, aditivo, de dominância e de contribuição gênica referentes aos grupos de ligação 1, 2 e 3. Observou-se que todos os efeitos foram diferentes dos simulados. É importante salientar que estes resultados, mesmo quando tiveram valores próximos dos valores simulados, foram influenciados pelos desvios padrão dos valores estimados, que foram pequenos. Outro ponto a ser considerado é a influência de um QTL nos efeitos do outro. Esta influência ficou mais evidente no grupo de ligação 2, em que o efeito referente ao primeiro QTL foi subestimado, quando comparado com os do grupo um, que explicavam a mesma porcentagem da característica (28%). Já o segundo QTL foi superestimado em relação ao terceiro grupo de ligação, que também explicava 5%.

Tabela 43. Médias, desvio padrão (entre parênteses) e valores do teste t para os efeitos das médias genótípicas (m) e aditivos (a) para as populações de 1000 indivíduos referentes aos grupos de ligação 1, 2 e 3.

QTLs	M			a		
	Estimado	Simulado	Valor t	Estimado	Simulado	Valor t
GL 1 QTL 1	107,06 (0,54)	100	92,70**	9,05 (0,75)	6,11	27,54**
GL 1 QTL 2	107,10 (0,58)	100	86,10**	9,08 (0,59)	6,11	35,50**
GL 2 QTL 1	107,62 (0,63)	100	85,19**	7,41 (0,65)	6,11	14,18**
GL 2 QTL 2	108,70 (0,72)	100	85,31**	6,13 (0,94)	2,58	26,76**
GL 3 QTL 1	109,15 (0,56)	100	116,21**	4,00 (0,55)	2,58	18,33**
GL 3 QTL 2	109,11 (0,59)	100	108,89**	3,89 (0,63)	2,58	14,65**

** - Significativo a 1% de probabilidade pelo teste t.

Tabela 44. Médias, desvio padrão (entre parênteses) e valores do teste t para os efeitos devidos ao desvio de dominância (d) e contribuição gênica (R^2) para as populações de 1000 indivíduos referentes aos grupos de ligação 1, 2 e 3.

QTLs	D			R^2		
	Estimado	Simulado	Valor t	Estimado	Simulado	Valor t
GL 1 QTL 1	7,54 (0,93)	6,11	10,90**	23,23 (3,16)	28	-10,69**
GL 1 QTL 2	7,41 (1,00)	6,11	9,10**	22,96 (2,08)	28	-17,15**
GL 2 QTL 1	6,52 (1,01)	6,11	2,87**	16,18 (2,10)	28	-39,85**
GL 2 QTL 2	4,21 (1,36)	2,58	10,15**	9,95 (2,47)	5	14,15**
GL 3 QTL 1	3,32 (0,90)	2,58	5,81**	4,60 (1,03)	5	-2,75**
GL 3 QTL 2	3,42 (0,95)	2,58	6,23**	4,55 (1,20)	5	-2,65**

** - Significativo a 1% de probabilidade pelo teste t.

3.2.2. Detecção de QTL por metodologias de intervalo

A metodologia proposta por Lander; Botstein (1989) flanqueia o QTL por meio de duas marcas. Para verificar a associação entre os marcadores e o QTL, pelo método de regressão, os três genótipos são codificados da mesma forma que são codificados para a análise de marca simples. Por essas análises são estimados os efeitos das médias, aditivos e de dominância e a contribuição gênica do QTL, além de posicioná-los no cromossomo, podendo ser utilizadas as distâncias em centimorgan, Haldane ou Kosambi (SCHUSTER; CRUZ, 2008).

A detecção e localização dos QTLs foram realizadas pela metodologia de intervalo simples. Na maioria das repetições, esta metodologia detectou os dois QTLs simulados, no entanto, houve repetições que detectaram mais de dois ou apenas um QTL e para esses casos foi utilizada a metodologia de intervalo composto. Os cofatores foram selecionados por meio

da regressão de *stepwise*, sendo utilizados os níveis de probabilidades de entrada e saída de variáveis no modelo iguais a 5% (Tabela 45). Verifica-se na Figura 13 que foram detectados mais de dois QTLs, quando utilizada a metodologia de intervalo simples e após a aplicação do intervalo composto detectou-se apenas dois QTLs (Figura 14).

Tabela 45. Resultado da análise de regressão por *stepwise* de uma característica quantitativa em função dos marcadores dos quatro grupos de ligação de um genoma simulado referente à uma repetição.

GL	Marca	R ² (%)	F	Prob(%)	Entrada	Saída
1	29	10,97	123,02	0,00	Confirmada	
1	35	12,64	72,12	0,00	Confirmada	Nenhuma
1	28	13,00	49,63	4,10	Confirmada	29
1	31	13,06	49,86	6,04	Não Confirmada	
2	50	4,14	43,10	0,00	Confirmada	
2	57	4,94	25,89	0,39	Confirmada	Nenhuma
2	61	5,17	18,09	11,89	Não Confirmada	
3	8	16,23	193,36	0,00	Confirmada	
3	15	20,46	128,23	0,00	Confirmada	Nenhuma
3	14	20,61	86,18	17,23	Não Confirmada	
4	82	0,23	2,28	13,13	Não Confirmada	

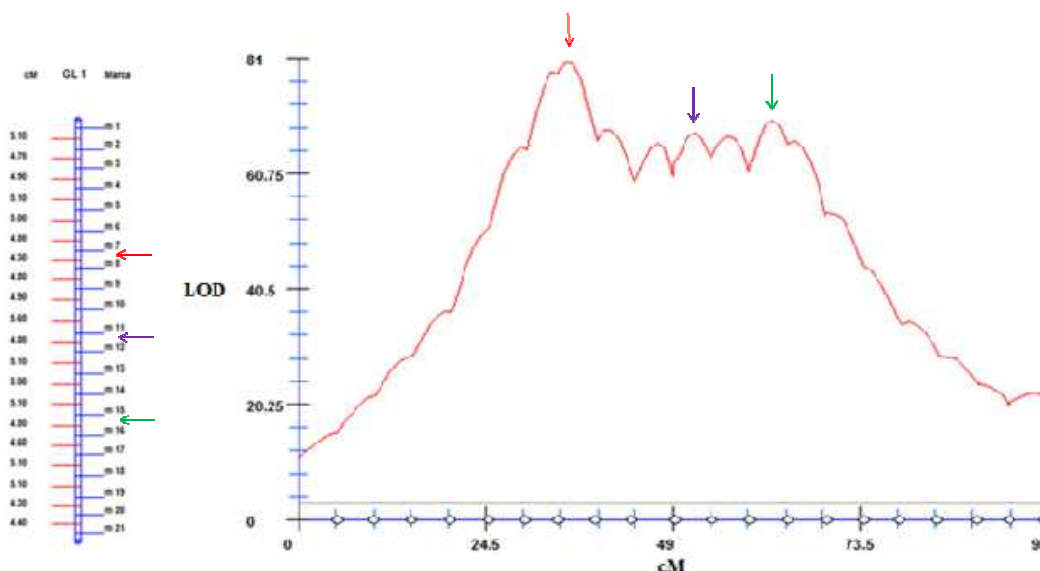


Figura 13. Análise de QTL, pelo método do intervalo simples, referente a uma população F2 obtida por simulação, em relação ao grupo de ligação 1. No eixo das ordenadas estão os valores de LOD e no eixo das abscissas os intervalos em cM.

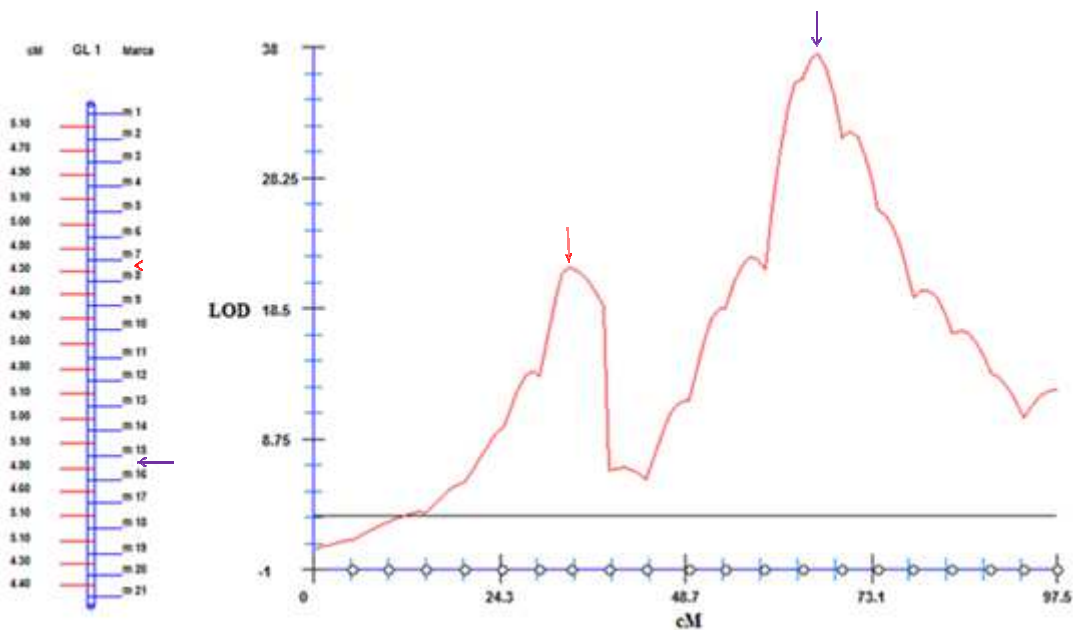


Figura 14. Análise de QTL, pelo método do intervalo composto, referente a uma população F2 obtida por simulação, em relação ao grupo de ligação 1. No eixo das ordenadas estão os valores de LOD e no eixo das abscissas os intervalos em cM.

Schuster; Cruz (2008) analisaram um conjunto de dados, obtidos por meio de simulação, pelo método de intervalo composto, que foram utilizados para fazer abordagem sobre a sua eficácia em relação ao método do intervalo simples e para subsidiar no procedimento de escolha de marcadores a serem empregados como cofatores. Neste contexto, foi realizada, inicialmente, a análise de regressão de *stepwise*, em que a característica quantitativa foi submetida a um modelo cujas variáveis independentes eram os marcadores dos respectivos grupos de ligação. Foram admitidos níveis de probabilidades de entrada e saída de variáveis no modelo iguais a 5%. Esses autores concluíram que o método de intervalo composto é eficiente na prevenção de detecção de QTL fantasma. Encontram-se na literatura trabalhos que utilizam a regressão de *stepwise* para selecionar cofatores (XU et al., 2011; BROWN et al., 2011; HENNING et al., 2011; LI et al., 2010; SU et al., 2010; XU et al., 2011; MORO, 2011).

Sabe-se que os grupos de ligação 1, 2 e 3 contêm dois QTLs cada, sendo o primeiro na posição 33,29 cM e o segundo à 66,59 cM. Foi identificado pelo menos um QTL em todos os grupos de ligação. Pela metodologia de intervalo simples, grupo de ligação 1, observou-se que 30% das repetições detectaram mais de dois QTLs e 2% somente um. Após a aplicação da metodologia de intervalo composto, em 8% das repetições o problema de detecção de falso positivo não foi resolvido, detectando mais de dois QTLs (Tabela 46).

Tabela 46. Número de repetições que detectaram mais de 2, somente 1 e nenhum QTL por grupo de ligação pelas metodologias de mapeamentos por intervalo simples (IS) e intervalo composto (IC), nos grupos de ligação 1, 2 e 3 das populações simuladas de 1000 indivíduos.

GL	Mais de 2 QTLs		2 QTLs		Somente 1 QTL		Nenhum QTL		Total
	IS	IC	IS	IC	IS	IC	IS	IC	
1	15 (30) ¹	4 (8)	34 (68)	46 (92)	1 (2)	0	0	0	50
2	4 (8)	1 (2)	21 (42)	41 (82)	25 (50)	8 (16)	0	0	50
3	24 (48)	9 (18)	23 (46)	40 (80)	3 (6)	1 (2)	0	0	50

¹Entre parênteses referem à percentagem.

No grupo de ligação 2 detectaram-se mais de dois QTLs em 8% das repetições e somente um em 50%, quando analisada pelo intervalo simples. Posteriormente a aplicação do intervalo composto, detectaram-se mais de dois e apenas um QTL em 2% e 16%, respectivamente. Referente ao grupo de ligação 3, quando se utilizou a metodologia de intervalo simples em 48% foram verificados mais de dois QTLs e em 6% foi encontrado somente um. Após o uso do intervalo composto 18% e 2% das repetições detectaram mais de dois QTLs e apenas um, respectivamente.

Considerando que no grupo de ligação 1 os dois QTLs explicavam 28% da variação genotípica, cada um, observou-se que a percentagem de detecção de mais de dois QTLs foi elevada, utilizando a metodologia de intervalo simples. No entanto, quando se aplicou a metodologia de intervalo composto esta percentagem diminuiu consideravelmente. Resultado semelhante ocorreu no grupo 3, no qual os dois QTLs explicam 5% da variação genotípica. Entretanto, no grupo de ligação três esta redução foi menor do que no grupo um, provavelmente devido ao fato dos QTLs explicarem menos a característica. No grupo 2, onde um QTL explica 28% e o outro 5%, a percentagem de repetições que identificaram apenas um QTL foi relativamente alta (50%), mas quando se aplicou o intervalo composto reduziu-se esta percentagem para 16%.

Observa-se que, quando os dois QTLs explicam a mesma percentagem da variação genotípica, há elevado percentual de detecção de QTLs fantasmas. Verificou-se também que, quando o efeito é menor, maior será a detecção deste QTL. No entanto, quando os efeitos dos QTLs são diferentes, no mesmo grupo de ligação, a detecção de apenas um QTL é maior. Neste caso, provavelmente, o QTL de maior efeito dificultou a detecção do QTL que explica menos a característica.

Quando se usa populações grandes, como a de 1000 indivíduos, e um QTL que explica grande parte da variação genotípica por grupo de ligação a metodologia de intervalo simples é eficiente. Caso este seja de pequeno efeito recomenda-se então a utilização da metodologia de intervalo composto, pois se pode evitar a detecção de QTL fantasma. Esta metodologia leva em consideração os cofatores, que são marcadores do mesmo ou de outros grupos de ligação. Desta forma, consegue-se minimizar a interferência do efeito de outros QTLs, tornando a detecção do QTL de interesse mais precisa (WU et al, 2007).

O mapeamento de intervalo composto aumenta a precisão das estimativas da posição e do efeito do QTL, utilizando marcadores cofatores que reduzem a variância residual provocada por outros QTLs fora do intervalo. Porém, uma questão que não tem solução simples é: quais os marcadores que devem ser utilizados na regressão como cofatores? Certamente os dois marcadores que estão flanqueando diretamente o intervalo devem ser incluídos. A escolha dos demais marcadores, que irão compor o termo $\sum_{k \neq i, i+1}^c \beta_k x_{kj}$, não é uma tarefa trivial. Esses cofatores podem ser selecionados utilizando a regressão de *stepwise*. É importante ressaltar que os efeitos dos QTLs podem ser reduzidos pela inclusão de marcadores ligados ao intervalo de interesse (CRUZ et al., 2009).

Observa-se na Tabela 47 que os primeiros QTLs proporcionam menores desvios padrão nos três grupos de ligação, portanto, as posições foram mais precisas que nos demais. No grupo 1, onde os dois QTLs explicavam 28% da característica, o desvio padrão foi menor. No grupo de ligação 3, onde se encontravam QTLs explicando 5% da variação genotípica, o desvio padrão do QTL na posição de 34,54 cM foi maior do que nos demais grupo e para o segundo QTL, o grupo 2 obteve o maior desvio padrão, entre os três grupos. Este maior desvio, provavelmente, foi devido a quanto cada QTL explica, pois para o segundo QTL, que explica 5% da característica, houve maior dificuldade para detectá-lo. Diante disto, pode-se afirmar que quando há presença de um QTL de maior e um de menor efeito no mesmo grupo de ligação, o de maior efeito pode dificultar a detecção do QTL de menor efeito.

Tabela 47. Médias e desvios padrão (entre parênteses) das posições dos QTLs das populações de tamanho 1000 indivíduos referentes aos grupos de ligação 1, 2 e 3.

QTLs	GL 1	GL 2	GL 3
QTL 1	35,86 (1,95)	33,76 (2,09)	34,54 (3,98)
QTL 2	63,66 (2,65)	62,46 (5,29)	63,73 (4,70)

Nas figuras 15 e 16 estão apresentadas as frequências das posições dos QTLs referentes aos grupos de ligação 1, 2 e 3. A maior amplitude foi de 57,17 cM a 67,75 cM obtida pelo posicionamento do segundo QTL no grupo de ligação 2, seguidos pelo segundo e primeiro QTL do terceiro grupo, primeiro do grupo 2 e primeiro QTL do grupo 1.

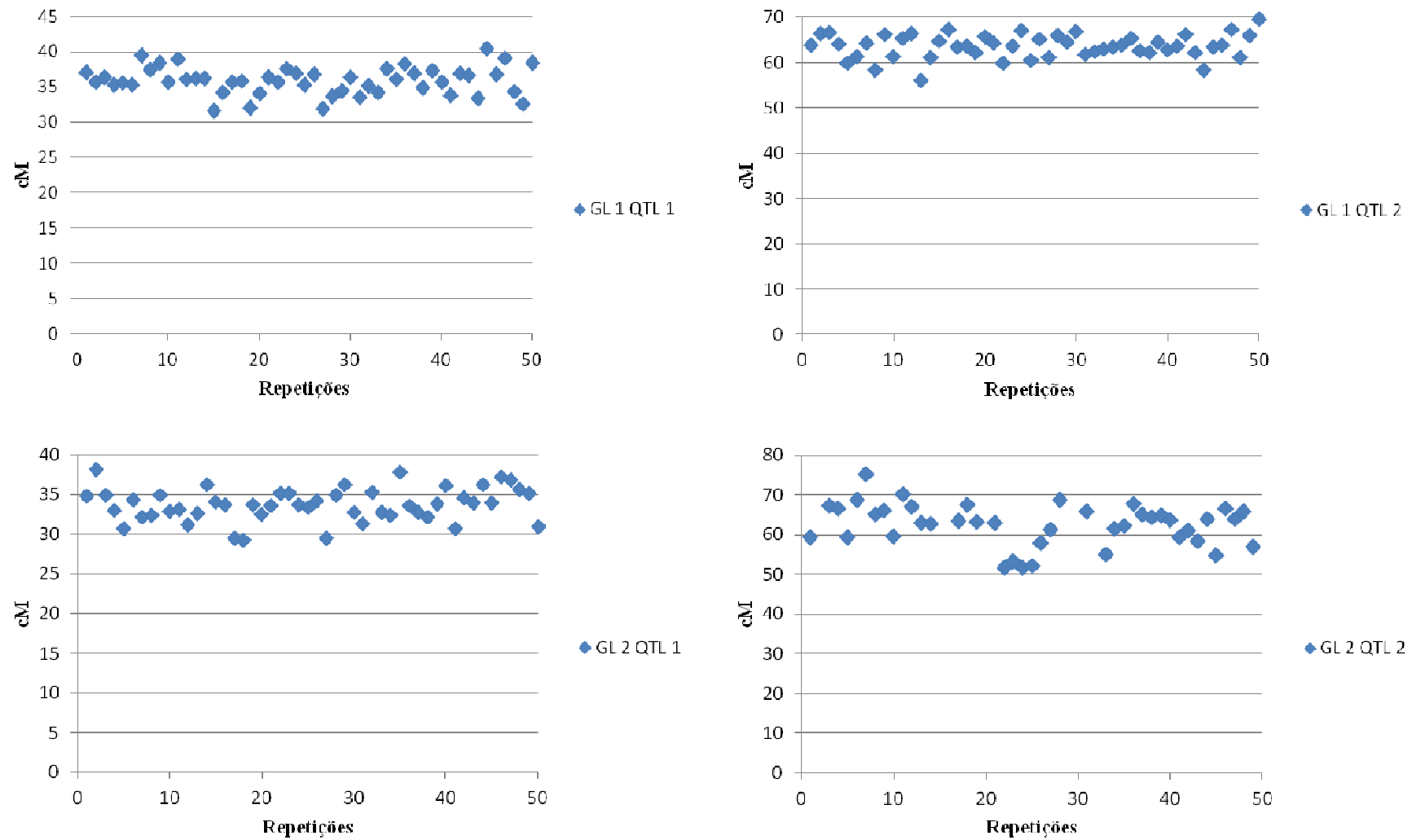


Figura 15. Frequências dos QTLs 1 e 2 nos grupos de ligação 1 e 2 obtidas nas populações simuladas de 1000 indivíduos. No eixo das ordenadas estão as posições e no eixo das abscissas as repetições.

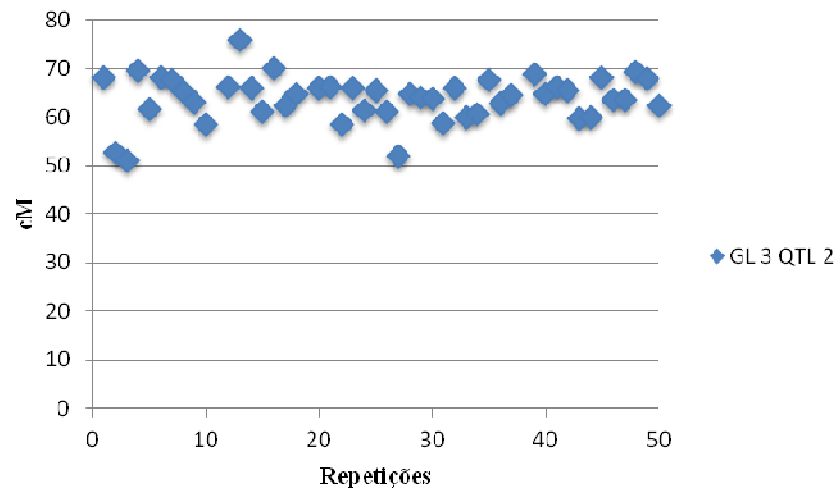
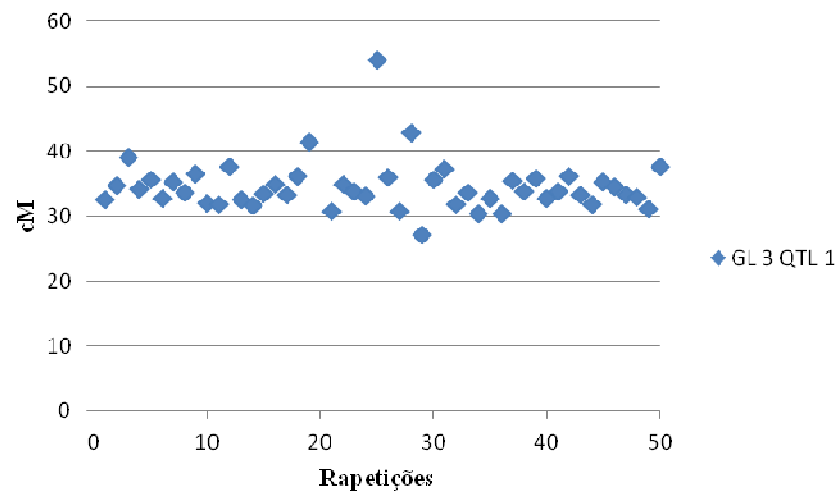


Figura 16. Frequências dos QTLs 1 e 2 no grupo de ligação 3 obtidas nas populações simuladas de 1000 indivíduos. No eixo das ordenadas estão as posições e no eixo das abscissas as réplicas.

Estão apresentadas, nas Tabelas 48, 49 e 50, as médias e os valores do teste t referentes aos efeitos das médias genotípicas, aditivo, de dominância, posição e contribuição gênica referentes aos grupos de ligação 1, 2 e 3.

Observa-se que as estimativas das médias genotípicas, efeito aditivo, dos desvios de dominância e posição dos QTLs foram diferentes dos valores simulados, em todos os grupos de ligação. Os valores referentes à contribuição gênica estimados foram não significativos, no grupo de ligação 3, para ambos QTLs e os demais foram diferentes dos simulados. No segundo grupo de ligação a contribuição gênica do primeiro QTL foi 10% a menos e o segundo mais de 4% a mais dos valores simulados. Isto mostra que tanto o primeiro QTL quanto o segundo QTL influenciam na detecção dos efeitos de ambos.

Tabela 48. Médias, desvio padrão (entre parênteses) e valores do teste t para os efeitos das médias genotípicas (m) e aditivos (a) para as populações de 1000 indivíduos referentes aos dois QTLs nos grupos de ligação 1, 2 e 3.

QTLs	m			A		
	Estimado	Simulado	Valor t	Estimado	Simulado	Valor t
GL 1 QTL 1	105,29 (2,89)	100	12,94**	8,94 (1,22)	6,11	16,45**
GL 1 QTL 2	105,19 (3,21)	100	11,43**	8,85 (1,41)	6,11	13,73**
GL 2 QTL 1	104,65 (3,55)	100	9,26**	7,69 (2,61)	6,11	4,28**
GL 2 QTL 2	104,89 (4,69)	100	7,38**	4,87 (2,04)	2,58	7,93**
GL 3 QTL 1	106,51 (4,12)	100	11,16**	3,93 (0,87)	2,58	11,02**
GL 3 QTL 2	106,23 (4,43)	100	9,94**	3,80 (0,88)	2,58	9,83**

** - significativo a 1% pelo teste t.

Tabela 49. Médias, desvio padrão (entre parênteses) e valores do teste t para os efeitos devidos ao desvio de dominância (d) e posicionamento do QTL (posição) para as populações de 1000 indivíduos referentes aos dois QTLs nos grupos de ligação 1, 2 e 3.

QTLs	d			Posição		
	Estimado	Simulado	Valor t	Estimado	Simulado	Valor t
GL 1 QTL 1	7,90 (0,96)	6,11	13,12**	35,86 (1,95)	33,29	9,31**
GL 1 QTL 2	7,76 (1,11)	6,11	10,48**	63,66 (2,66)	66,59	-7,80**
GL 2 QTL 1	6,97 (1,49)	6,11	4,08**	33,76 (2,09)	33,29	1,59 ^{ns}
GL 2 QTL 2	4,38 (0,98)	2,58	12,92**	62,46 (5,29)	66,59	-5,52**
GL 3 QTL 1	3,47 (1,04)	2,58	6,06**	34,54 (3,98)	33,29	2,22*
GL 3 QTL 2	3,62 (1,01)	2,58	7,25**	63,73 (4,70)	66,59	-4,30**

^{ns}, * e** - não significativo, significativo a 5% e a 1%, respectivamente, pelo teste t.

Tabela 50. Médias, desvio padrão (entre parênteses) e valores do teste t para o efeito da contribuição gênica (R^2) para as populações de 1000 indivíduos referentes aos dois QTLs nos grupos de ligação 1, 2 e 3.

QTLs	Ação		
	Estimado	Simulado	Valor t
GL 1 QTL 1	32,73 (10,96)	28	3,05**
GL 1 QTL 2	33,10 (8,68)	28	4,15**
GL 2 QTL 1	18,11 (7,48)	28	-9,35**
GL 2 QTL 2	9,33 (4,64)	5	6,60**
GL 3 QTL 1	4,71 (1,82)	5	-1,13 ^{ns}
GL 3 QTL 2	4,95 (1,84)	5	-0,19 ^{ns}

^{ns}, * e ** - não significativo, significativo a 5% e a 1%, respectivamente, pelo teste t.

Rodrigues et al. (2010), com o objetivo de detectar e mapear QTL que afetam os conteúdos de proteína e óleo em soja, utilizaram 48 marcadores microssatélites (SSR) polimórficos e 207 famílias das progênes. Foram detectados quatro QTLs associados ao conteúdo de proteína e três QTLs associados ao de óleo explicando de 6,24 a 18,94% e 17,26 a 25,93%, respectivamente, para os conteúdos de proteína e óleo. Para a detecção destes QTLs foram utilizadas as metodologias de intervalo simples e composto.

Com o intuito de detectar QTLs ligados à época de floração da *Arabidopsis thaliana*, Salomé et al. (2011) estudaram 17 populações F₂. A detecção destes QTLs foi realizada por meio das metodologias de intervalo simples e composto, sendo que este visou a não detecção do falso positivo. Chakraborty; Zeng (2011) utilizaram essas metodologias para estudar e mapear QTLs ligados aos dias de florescimento de arroz e foram detectados cinco QTLs explicando a característica.

3.3. Considerações referentes à detecção de QTLs fantasmas

Quando se busca encontrar QTL, primeiramente precisa-se definir se o objetivo é apenas detectar ou detectar e posicionar. Mesmo quando o interesse for apenas a detecção pode ocorrer mais de um QTL no mesmo grupo de ligação e os efeitos de um pode interferir no do outro, de forma a super ou subestimar seus efeitos. Além disso, essas interferências podem ocasionar o surgimento de QTL fantasma (SILVA; VENCOVSKY, 2002; SCHUSTER; CRUZ, 2008; JANGARELLI et al., 2011).

Como foi visto, a metodologia da marca simples, por meio de regressão linear, é eficiente em detectar marcadores ligados ao QTL e quantificar seus efeitos. No entanto, quando há presença de mais de um QTL este método não detecta se há ou não QTL fantasma ou se um suprime o efeito do outro.

Observou-se, no Capítulo 1, que a metodologia de intervalo simples é eficiente para estudo de detecção e localização de QTL, principalmente quando se utiliza populações grandes, como a de 1000 indivíduos. Outro fato a ser observado é que em casos de problema na detecção, o uso de metodologia de intervalo composto é recomendado.

A metodologia de intervalo simples foi eficiente na detecção dos QTLs, entretanto, foram observados que na presença de dois QTLs que explicam a mesma porcentagem da característica a detecção de QTL fantasma foi elevada. No entanto, quando foram submetidos à análise de intervalo composto, a porcentagem de QTL fantasma detectada caiu consideravelmente. Verificou-se que, quanto menos os QTLs explicavam, maior foi a porcentagem de QTL fantasma. Observou-se também que, quando os QTLs explicavam de maneira diferente a variação genotípica, houve uma tendência de o QTL de maior efeito suprimir o de menor efeito. Este problema também foi minimizado com a utilização da metodologia de intervalo composto.

Diante disto, pode-se afirmar que a metodologia de intervalo composto é eficiente na detecção de QTL visando a não detecção de fantasmas e que um QTL suprime o efeito do QTL de menor efeito. Corroborando tais afirmações, são encontrados na literatura trabalhos que utilizam intervalo composto visando solucionar esses problemas (CHAKRABORTY; ZENG, 2011; XU et al., 2011; DING et al., 2011; LIU et al., 2011; LU et al., 2011; YUAN et al., 2011). Trabalhos com simulações, com o intuito de analisar a eficiência dessas metodologias na prevenção desses problemas, não foram encontrados na literatura.

4. CONCLUSÕES

Para fins de mapeamento, conclui-se que:

O tamanho populacional de 1000 indivíduos recupera os grupos de ligação e mantém as ordem dos marcadores.

Para fins de detecção e posicionamento de QTL, conclui-se que:

As metodologias de intervalo simples e composto são eficientes na detecção e no posicionamento de QTL;

A metodologia de intervalo composto melhorou a não detecção de QTL fantasma e que o QTL de maior efeito suprime o de menor efeito, quando comparado à metodologia de intervalo simples;

Nos grupos de ligação com dois QTLs com o mesmo efeito há maior tendência para surgir QTL fantasma;

Em grupo de ligação com dois QTLs de pequeno efeito a eficiência na prevenção de QTL fantasma é menor que nos de maior efeito;

Em grupo de ligação com dois QTLs contribuindo diferentemente para a variação genotípica, o de maior pode suprimir o de menor contribuição gênica.

Para fins de efeitos do QTL, conclui-se que:

Quando se tem dois QTLs no mesmo grupo de ligação os efeitos podem ser super ou subestimados.

5. REFERÊNCIAS BIBLIOGRÁFICAS

BHERING, L.L.; CRUZ, C.D. Tamanho de população ideal para mapeamento genético em famílias de irmãos completos. **Pesquisa agropecuária brasileira**, v.43, p.379-385, 2008.

BROWN, A. A.; RICHARDSON, S.; WHITTAKER, T. Application of the lasso to expression Quantitative Trait Loci mapping. **Statistical Application in Genetics and Molecular Biology**, v.10, n.1, 2011. DOI: 10.2202/1544-6115.1606

CAIXETA, E. T.; OLIVEIRA, A. C. B.; BRITO, G. G.; SAKIYAMA, N. S. Tipos de Marcadores Moleculares. In: BORÉM, A.; CAIXETA, E. T. **Marcadores moleculares**. 2ª Ed., ed. UFV, Viçosa, p.483-532, 2009.

CHAKRABORTY, S.; ZENG, Z. B. QTL mapping for Days to Flowering under Drought Condition in Rice (*Oryza sativa* L.) Genome. **Not Bot Hort Agrobot Cluj**, v.39, n.1, p.58-63, 2011.

CRUZ, C. D.; BHERING, L. L.; GOD, P. I. V. G. Mapeamento de QTLs em populações derivadas de cruzamentos controlados. In: BORÉM, A.; CAIXETA, E. T. **Marcadores moleculares**. 2ª Ed., ed. UFV, Viçosa, p.483-532, 2009.

CRUZ, C. D. **Programa para análise de dados moleculares e quantitativos – GQMOL - Versão 2008.6.1**. Viçosa: UFV, 2008.

CRUZ, E. M. **Efeito da saturação e do tamanho da populações F2 e de retrocruzamento sobre a acurácia do mapeamento genético**. Viçosa, MG: UFV, 2006. (Tese de Doutorado em Genética e Melhoramento). Universidade Federal de Viçosa, Viçosa.

DING, J. Q.; MA, J. L.; ZHANG, C. R.; DONG, H. F.; XI, Z. Y.; XIA, Z. L.; WU, J. Y. QTL mapping for test weight by using F_{2:3} population in maize. **Journal of Genetics**, v.90, n.1, p.75-80, 2011.

HENNING, J. A.; TOWNSEND, M. S.; GENT, D. H.; BASSIL, N.; MATTHEWS, P.; BUCK, E.; BEATSON, R. QTL mapping of powdery mildew susceptibility in hop (*Humulus lupulus* L.). **Euphytica**, v.180, p.411-420, 2011. DOI: 10.1007/s10681-011-0403-4

JANGARELLI, M.; EUCLYDES, R. F.; CECON, P. R. Níveis de significância na identificação de marcadores moleculares no mapeamento genômico. **Revista Brasileira de Zootecnia**, v.40, n.2, p.308-313, 2011.

LANDERS, S.; BOTSTEIN, D. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. **Genetics**, v.121, n.1, p.185-199, 1989.

LAPERUTA, L. D. C. **Estudo de uma população segregante (F₁) de maracujá doce: enriquecimento do mapa de ligação e mapeamento de QTL para produção e qualidade de frutos**. Piracicaba, SP: ESALQ, 2011. 134p. (Tese de doutorado em Ciências). Universidade de São Paulo, Piracicaba.

LI, H.; HEARNE, S.; BÄNZIGER, M.; LI, Z.; WANG, J. Statistical properties of QTL linkage mapping in bilateral genetic population. **Heredity**, v.105, p.257-267, 2010. DOI: 10.1038/hdy.2010.56

LIU, R.; WANG, B.; GUO, W.; WANG, L.; ZHANG, T. Differential gene expression and associated QTL mapping for cotton yield based on a cDNA-AFLP transcriptome map in an immortalized F₂. **Theor Appl Genet**, 2011. DOI: 10.1007/s00122-011-1597-5

LU, B.; XIE, K.; YANG, C.; ZHANG, L.; WU, T.; LI, L.; LIU, X.; JIANG, L.; WAN, J. Efficient QTL detection for heading date in bechcross inbred line and F₂ population derived from the same rice cross. **African Journal of Agriculture Research**, v.6, n.10, p.2372-2378, 2011.

MORO, G. V. **Uso de seleção genômica e fenotípica de linhagens para a predição de testecrosses em milho**. Piracicabe, SP: ESALQ, 2011. 116p. (Tese de Doutorado em Ciências). Universidade de São Paulo, Piracicaba.

OLIVER, R. E.; JELLEN, E. N.; LADIZINSKY, G.; KORAL, A. B.; KILIAN, A.; BEARD, J. L.; DUMLUPINAR, Z.; WISNIEWSKI-MOREHEAD, E.; SVEDIN, E.; COON, M.; REDMAN, R. R.; MAUGHAN, P. J.; OBERT, D. E.; JACKSON, E. W. New Diversity Arrays Technology (DARt) markers for tetraploid oat (*Avena magna* Murphy et Terrell) provide the first complete oat linkage map and markers linked to domestication genes from hexaploid *A. sativa* L. **Theoretical and Applied Genetics**, v.123, n.7, p.1159-1171, 2011. DOI: 10.1007/s00122-011-1656-y

RODRIGUES, J. I. S.; MIRANDA, F. D.; FERREIRA, A.; BORGES, L. L.; FERREIRA, M. F. S.; GOOD-GOD, P. I. V.; PIOVESAN, N. D.; BARROS, E. G.; CRUZ, C. D.; MOREIRA, M. A. Mapeamento de QTL para conteúdo de proteína e óleo em soja. **Pesquisa Agropecuária Brasileira**, v.45, n.5, p.472-480, 2010.

SALOMÉ, P. A.; BOMBLIES, K.; LAITINEN, R. A. E.; YANT, L.; MOTT, R.; WEIGEL, D. Genetic architecture of flowering time variation in *Arabidopsis thaliana*. **Genetics**, 2011. DOI: 10.1534/genetics.111.126607

SCHUSTER, I.; CRUZ, C.D. **Estatística genômica aplicada a populações derivadas de cruzamento controlados**. 2º Ed. Viçosa, 2008. 568p.

SILVA, H. D.; VENCOVSKY, R. Poder de detecção de “Quantitative Trait Loci”, da análise de marca simples e da regressão linear múltipla. **Scentia Agricola**, v.59, n.4, p.755-762, 2002.

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

SU, C. F.; LU, W. G.; ZHAO, T. J.; GAI, J. Y. Verification and fine-mapping of QTLs conferring days to flowering in soybean using residual heterozygous lines. **Chinese Science Bulletin**, v.55, n.6, p.499-508, 2010. DOI: 10.1007/s11434-010-0032-7

TOMÉ, L. G. O. **Análise biométrica e mapeamento de QTLs para tolerância à seca em milho**. Viçosa, MG: UFV, 2005. 148p. (Tese de Doutorado em Biologia Celular). Universidade Federal de Viçosa, Viçosa.

XU, M. H.; WEI, C. S.; TANG, X. T.; ZHU, Z. H.; SIMA, Y. F.; LOU, X. Y. A new mapping method for quantitative trait loci of silkworm. **Genetics**, 2011. DOI: 10.1186/1471-2156-12-19

WU, R.; MA, C. X.; CASALLA, G. **Statistical Genetics of Quantitative Traits: Linkage, Maps, and QTL**. New York, NY: Springer, 2007. 365 p.

YUAN, J.; WANG, B.; WU, L. Quantitative trait loci (QTL) mapping for inflorescence length traits in *Lablab purpureus* (L.) sweet. **African Journal of Biotechnology**, v.10, n.18, p.3558-3566, 2011.

ZENG, Z. B. Theoretical basis of precision mapping of quantitative trait loci. **Proceedings of the National Academy of Science United States of America**, v.90, p.10972-10976, 1993.

APÊNDICES

Apêndice 1. Análise da metodologia da marca simples por meio de regressão linear para os 84 marcadores, referente à uma repetição.

Marca	Prob(F)	B0	B1	Prob(t)	B2	Prob(t)	R ² (%)
m 1	0,00**	110,70	3,68	0,00**	1,66	0,08 ^{ns}	3,12
m 2	0,00**	109,94	4,09	0,00**	3,10	0,00**	4,48
m 3	0,00**	109,33	4,75	0,00**	4,13	0,00**	6,30
m 4	0,00**	108,87	5,37	0,00**	4,91	0,00**	8,14
m 5	0,00**	108,66	6,12	0,00**	5,17	0,00**	9,97
m 6	0,00**	108,66	7,07	0,00**	5,23	0,00**	12,69
m 7	0,00**	107,87	7,91	0,00**	6,81	0,00**	17,27
m 8	0,00**	107,64	8,86	0,00**	7,35	0,00**	21,50
m 9	0,00**	107,87	8,50	0,00**	6,95	0,00**	19,77
m 10	0,00**	107,75	8,20	0,00**	6,97	0,00**	18,40
m 11	0,00**	107,58	8,31	0,00**	7,13	0,00**	18,63
m 12	0,00**	107,30	9,01	0,00**	7,61	0,00**	21,62
m 13	0,00**	107,12	8,97	0,00**	7,99	0,00**	22,14
m 14	0,00**	106,62	9,32	0,00**	8,46	0,00**	23,06
m 15	0,00**	107,16	8,90	0,00**	7,42	0,00**	19,94
m 16	0,00**	107,83	7,81	0,00**	6,20	0,00**	14,80
m 17	0,00**	108,17	6,50	0,00**	5,75	0,00**	11,02
m 18	0,00**	109,12	5,36	0,00**	4,22	0,00**	7,18
m 19	0,00**	109,93	4,92	0,00**	2,86	0,00**	5,57
m 20	0,00**	110,10	4,54	0,00**	2,56	0,00**	4,71
m 21	0,00**	110,55	3,64	0,00**	1,78	0,06 ^{ns}	2,98
m 22	0,00**	111,11	3,99	0,00**	1,08	0,25 ^{ns}	3,50
m 23	0,00**	110,39	4,51	0,00**	2,49	0,00**	5,02
m 24	0,00**	109,96	5,09	0,00**	3,30	0,00**	6,67
m 25	0,00**	108,97	5,74	0,00**	5,03	0,00**	9,51
m 26	0,00**	108,39	6,94	0,00**	6,17	0,00**	14,02
m 27	0,00**	107,94	7,39	0,00**	7,24	0,00**	17,21
m 28	0,00**	107,81	7,98	0,00**	7,49	0,00**	19,49
m 29	0,00**	107,62	8,43	0,00**	7,76	0,00**	21,25
m 30	0,00**	108,30	7,42	0,00**	6,51	0,00**	16,18
m 31	0,00**	108,69	6,77	0,00**	5,75	0,00**	13,28
m 32	0,00**	108,94	6,38	0,00**	5,29	0,00**	11,66
m 33	0,00**	109,01	5,72	0,00**	4,99	0,00**	9,48
m 34	0,00**	109,31	6,21	0,00**	4,30	0,00**	9,83
m 35	0,00**	109,24	6,45	0,00**	4,50	0,00**	10,68
m 36	0,00**	110,05	5,95	0,00**	3,01	0,00**	8,28
m 37	0,00**	110,20	5,71	0,00**	2,64	0,00**	7,46
m 38	0,00**	110,55	5,22	0,00**	1,87	0,04*	5,99
m 39	0,00**	110,77	4,95	0,00**	1,51	0,10 ^{ns}	5,37
m 40	0,00**	111,49	4,53	0,00**	0,11	0,90 ^{ns}	4,32
m 41	0,00**	111,68	4,02	0,00**	-0,27	0,77 ^{ns}	3,48
m 42	0,00**	111,13	3,26	0,00**	0,89	0,64 ^{ns}	2,45
m 43	0,13 ^{ns}	111,21	1,30	0,05 ^{ns}	0,63	0,52 ^{ns}	0,40

Apêndice 1. Continuação.

m 44	0,03*	111,17	1,70	0,01*	0,71	0,53 ^{ns}	0,67
m 45	0,03*	111,17	1,70	0,01*	0,71	0,53 ^{ns}	0,69
m 46	0,01**	110,95	1,93	0,00**	1,13	0,24 ^{ns}	0,93
m 47	0,01*	110,89	1,85	0,00**	1,20	0,21 ^{ns}	0,86
m 48	0,00**	110,27	2,37	0,00**	2,43	0,01*	1,77
m 49	0,00**	110,04	2,84	0,00**	2,77	0,00**	2,40
m 50	0,00**	109,61	3,38	0,00**	3,48	0,00**	3,43
m 51	0,00**	109,83	3,19	0,00**	3,20	0,00**	3,11
m 52	0,00**	110,25	3,46	0,00**	2,38	0,01*	3,03
m 53	0,00**	110,35	3,22	0,00**	2,32	0,01*	2,75
m 54	0,00**	110,38	3,26	0,00**	2,25	0,01*	2,80
m 55	0,00**	110,35	3,14	0,00**	2,47	0,01*	2,86
m 56	0,00**	110,11	3,14	0,00**	2,95	0,00**	3,10
m 57	0,00**	110,38	2,66	0,00**	2,50	0,00**	2,27
m 58	0,00**	110,26	2,16	0,00**	2,70	0,00**	1,83
m 59	0,00**	109,88	1,33	0,04*	3,43	0,00**	1,66
m 60	0,04*	110,76	1,28	0,05 ^{ns}	1,60	0,09 ^{ns}	0,64
m 61	0,06 ^{ns}	110,86	1,25	0,06 ^{ns}	1,39	0,14 ^{ns}	0,55
m 62	0,24 ^{ns}	111,02	0,85	0,21 ^{ns}	1,06	0,27 ^{ns}	0,28
m 63	0,11 ^{ns}	110,65	0,71	0,30 ^{ns}	1,74	0,06 ^{ns}	0,44
m 64	0,02*	110,75	1,44	0,03*	1,65	0,08 ^{ns}	0,76
m 65	0,04*	110,93	1,39	0,03*	1,32	0,16 ^{ns}	0,63
m 66	0,12 ^{ns}	110,78	0,83	0,21 ^{ns}	1,60	0,09 ^{ns}	0,43
m 67	0,08 ^{ns}	110,61	0,63	0,65 ^{ns}	1,93	0,04*	0,48
m 68	0,02*	110,31	0,60	0,62 ^{ns}	2,60	0,00**	0,80
m 69	0,04*	110,43	0,48	0,52 ^{ns}	2,29	0,01*	0,61
m 70	0,02*	110,34	0,70	0,29 ^{ns}	2,52	0,00**	0,78
m 71	0,04*	110,42	0,51	0,55 ^{ns}	2,35	0,01*	0,64
m 72	0,27 ^{ns}	110,83	0,36	0,59 ^{ns}	1,48	0,12 ^{ns}	0,26
m 73	100,00 ^{ns}	111,04	0,15	0,80 ^{ns}	1,00	0,30 ^{ns}	0,11
m 74	100,00 ^{ns}	111,18	0,26	0,70 ^{ns}	0,69	0,51 ^{ns}	0,06
m 75	100,00 ^{ns}	111,43	0,17	0,80 ^{ns}	0,18	,848 ^{ns}	0,00
m 76	100,00 ^{ns}	111,73	0,19	0,77 ^{ns}	-0,41	0,67 ^{ns}	0,02
m 77	100,00 ^{ns}	111,77	0,13	0,84 ^{ns}	-0,48	0,62 ^{ns}	0,02
m 78	100,00 ^{ns}	111,49	0,37	0,59 ^{ns}	0,05	0,95 ^{ns}	0,02
m 79	100,00 ^{ns}	111,31	0,64	0,64 ^{ns}	0,41	0,67 ^{ns}	0,10
m 80	100,00 ^{ns}	111,38	0,66	0,65 ^{ns}	0,29	0,76 ^{ns}	0,10
m 81	100,00 ^{ns}	111,50	0,52	0,54 ^{ns}	0,04	0,96 ^{ns}	0,05
m 82	100,00 ^{ns}	110,89	0,26	0,69 ^{ns}	1,26	0,19 ^{ns}	0,18
m 83	100,00 ^{ns}	111,22	0,50	0,52 ^{ns}	0,58	0,55 ^{ns}	0,08
m 84	100,00 ^{ns}	110,99	0,23	0,73 ^{ns}	1,08	0,26 ^{ns}	0,13

^{ns}, * e ** - Não significativos, significativos a 5% e 1%, respectivamente.

Apêndice 2. Análise da metodologia da marca simples por meio de regressão linear para os 84 marcadores, referente à uma repetição.

Marca	Prob(F)	B0	B1	Prob(t)	B2	Prob(t)	R ² (%)
m 1	0,00**	111,28	3,71	0,00**	0,42	0,65 ^{ns}	3,19
m 2	0,00**	111,55	3,98	0,00**	-0,15	0,86 ^{ns}	3,59
m 3	0,00**	111,31	4,20	0,00**	0,24	0,79 ^{ns}	4,00
m 4	0,00**	110,64	5,07	0,00**	1,46	0,10 ^{ns}	5,97
m 5	0,00**	110,25	5,92	0,00**	2,09	0,018*	8,35
m 6	0,00**	109,21	7,04	0,00**	3,92	0,00**	12,55
m 7	0,00**	107,96	7,77	0,00**	6,06	0,00**	16,61
m 8	0,00**	107,96	8,36	0,00**	6,02	0,00**	18,78
m 9	0,00**	108,28	8,13	0,00**	5,52	0,00**	17,62
m 10	0,00**	108,47	8,25	0,00**	4,94	0,00**	16,70
m 11	0,00**	108,90	8,18	0,00**	4,25	0,00**	16,20
m 12	0,00**	108,81	8,31	0,00**	4,84	0,00**	18,04
m 13	0,00**	108,73	8,84	0,00**	5,21	0,00**	20,57
m 14	0,00**	108,30	8,52	0,00**	5,81	0,00**	19,62
m 15	0,00**	108,20	7,74	0,00**	6,05	0,00**	17,05
m 16	0,00**	108,59	6,88	0,00**	5,59	0,00**	14,28
m 17	0,00**	109,06	6,79	0,00**	4,48	0,00**	12,45
m 18	0,00**	109,39	5,97	0,00**	3,99	0,00**	9,80
m 19	0,00**	110,27	4,30	0,00**	2,37	0,01**	4,87
m 20	0,00**	110,12	4,62	0,00**	2,64	0,00**	5,72
m 21	0,00**	110,50	3,94	0,00**	2,08	0,02*	4,30
m 22	0,00**	110,73	4,44	0,00**	1,32	0,14 ^{ns}	4,60
m 23	0,00**	111,02	4,68	0,00**	0,80	0,62 ^{ns}	5,00
m 24	0,00**	110,81	5,16	0,00**	1,30	0,14 ^{ns}	6,36
m 25	0,00**	109,91	5,93	0,00**	3,09	0,00**	9,25
m 26	0,00**	109,65	6,37	0,00**	3,79	0,00**	11,35
m 27	0,00**	109,19	7,14	0,00**	4,85	0,00**	15,01
m 28	0,00**	108,49	7,58	0,00**	6,11	0,00**	17,80
m 29	0,00**	108,58	8,17	0,00**	6,16	0,00**	20,36
m 30	0,00**	108,90	7,48	0,00**	5,45	0,00**	16,54
m 31	0,00**	109,48	7,04	0,00**	4,27	0,00**	13,66
m 32	0,00**	109,80	6,46	0,00**	3,62	0,00**	11,28
m 33	0,00**	110,26	5,97	0,00**	2,70	0,00**	9,10
m 34	0,00**	110,27	5,51	0,00**	2,67	0,00**	7,87
m 35	0,00**	110,10	5,34	0,00**	3,05	0,00**	7,76
m 36	0,00**	110,20	5,17	0,00**	2,74	0,00**	7,01
m 37	0,00**	110,10	4,13	0,00**	2,81	0,00**	4,85
m 38	0,00**	110,49	3,68	0,00**	2,08	0,02*	3,63
m 39	0,00**	109,95	3,13	0,00**	3,11	0,00**	3,35
m 40	0,00**	110,47	2,45	0,00**	2,13	0,02*	1,94
m 41	0,00**	110,59	2,49	0,00**	1,89	0,04*	1,88
m 42	0,00**	111,26	2,10	0,00**	0,51	0,59 ^{ns}	1,08
m 43	0,03*	111,65	1,76	0,01**	-0,33	0,72 ^{ns}	0,72

Apêndice 2. Continuação.

m 44	0,02*	111,15	1,82	0,01**	0,59	0,53 ^{ns}	0,74
m 45	0,00**	111,08	2,27	0,00**	0,74	0,57 ^{ns}	1,17
m 46	0,00**	110,57	2,08	0,00**	1,75	0,05 ^{ns}	1,27
m 47	0,00**	110,13	2,25	0,00**	2,61	0,00**	1,86
m 48	0,00**	109,95	2,76	0,00**	2,96	0,00**	2,65
m 49	0,00**	109,37	2,91	0,00**	4,09	0,00**	3,73
m 50	0,00**	109,29	3,04	0,00**	4,40	0,00**	4,34
m 51	0,00**	109,36	3,12	0,00**	4,33	0,00**	4,43
m 52	0,00**	109,15	3,26	0,00**	4,72	0,00**	5,06
m 53	0,00**	109,62	3,31	0,00**	3,95	0,00**	4,46
m 54	0,00**	109,83	3,38	0,00**	3,50	0,00**	4,18
m 55	0,00**	109,75	3,38	0,00**	3,62	0,00**	4,23
m 56	0,00**	109,79	3,18	0,00**	3,46	0,00**	3,71
m 57	0,00**	109,32	2,88	0,00**	4,39	0,00**	4,15
m 58	0,00**	110,22	3,26	0,00**	2,57	0,00**	3,19
m 59	0,00**	110,69	3,81	0,00**	1,59	0,07 ^{ns}	3,62
m 60	0,00**	111,04	3,37	0,00**	0,92	0,32 ^{ns}	2,64
m 61	0,00**	110,93	3,26	0,00**	1,13	0,21 ^{ns}	2,50
m 62	0,00**	110,81	3,42	0,00**	1,36	0,13 ^{ns}	2,78
m 63	0,00**	110,64	3,37	0,00**	1,72	0,06 ^{ns}	2,84
m 64	100,00 ^{ns}	111,81	0,11	0,86 ^{ns}	-0,55	0,56 ^{ns}	0,04
m 65	0,24 ^{ns}	112,34	0,23	0,73 ^{ns}	-1,55	0,09 ^{ns}	0,29
m 66	0,28 ^{ns}	112,28	0,19	0,77 ^{ns}	-1,47	0,11 ^{ns}	0,26
m 67	100,00 ^{ns}	111,70	0,04	0,95 ^{ns}	-0,31	0,73 ^{ns}	0,01
m 68	100,00 ^{ns}	111,53	0,09	0,88 ^{ns}	0,02	0,98 ^{ns}	0,00
m 69	100,00 ^{ns}	111,27	0,17	0,79 ^{ns}	0,54	0,56 ^{ns}	0,04
m 70	100,00 ^{ns}	111,36	0,32	0,64 ^{ns}	0,37	0,69 ^{ns}	0,04
m 71	100,00 ^{ns}	111,23	0,34	0,62 ^{ns}	0,64	0,50 ^{ns}	0,08
m 72	100,00 ^{ns}	111,26	0,12	0,85 ^{ns}	0,57	0,55 ^{ns}	0,04
m 73	100,00 ^{ns}	110,96	-0,14	0,82 ^{ns}	1,12	0,23 ^{ns}	0,14
m 74	100,00 ^{ns}	111,55	0,18	0,79 ^{ns}	-0,015	0,98 ^{ns}	0,01
m 75	100,00 ^{ns}	111,42	-0,09	0,89 ^{ns}	0,24	0,79 ^{ns}	0,01
m 76	100,00 ^{ns}	111,11	-0,72	0,28 ^{ns}	0,79	0,60 ^{ns}	0,18
m 77	100,00 ^{ns}	111,13	-0,68	0,30 ^{ns}	0,75	0,58 ^{ns}	0,16
m 78	0,12 ^{ns}	110,94	-1,14	0,08 ^{ns}	1,113	0,23 ^{ns}	0,43
m 79	0,04*	110,76	-1,36	0,04*	1,42	0,12 ^{ns}	0,62
m 80	0,14 ^{ns}	111,02	-1,17	0,08 ^{ns}	0,95	0,30 ^{ns}	0,40
m 81	0,32 ^{ns}	111,14	-0,90	0,18 ^{ns}	0,71	0,55 ^{ns}	0,23
m 82	0,31 ^{ns}	111,33	-0,99	0,13 ^{ns}	0,37	0,69 ^{ns}	0,23
m 83	100,00 ^{ns}	111,25	-0,35	0,60 ^{ns}	0,55	0,56 ^{ns}	0,06
m 84	100,00 ^{ns}	111,32	-0,33	0,62 ^{ns}	0,42	0,66 ^{ns}	0,04

^{ns}, * e ** - Não significativos, significativos a 5% e 1%, respectivamente.

Apêndice 3. Análise da metodologia da marca simples por meio de regressão linear para os 84 marcadores, referente à uma repetição.

Marca	Prob(F)	B0	B1	Prob(t)	B2	Prob(t)	R ² (%)
m 1	0,00**	110,38	3,71	0,00**	1,87	0,04*	3,29
m 2	0,00**	109,89	4,23	0,00**	2,85	0,00**	4,59
m 3	0,00**	109,45	5,18	0,00**	3,65	0,00**	7,02
m 4	0,00**	108,43	6,24	0,00**	5,65	0,00**	11,34
m 5	0,00**	107,95	6,80	0,00**	6,47	0,00**	13,69
m 6	0,00**	107,81	7,74	0,00**	6,70	0,00**	16,81
m 7	0,00**	107,29	7,99	0,00**	7,83	0,00**	19,48
m 8	0,00**	107,04	8,67	0,00**	8,12	0,00**	21,98
m 9	0,00**	106,98	9,08	0,00**	7,84	0,00**	22,11
m 10	0,00**	107,25	9,24	0,00**	7,35	0,00**	21,90
m 11	0,00**	106,91	9,25	0,00**	8,02	0,00**	23,13
m 12	0,00**	107,31	9,32	0,00**	7,65	0,00**	23,73
m 13	0,00**	107,27	8,85	0,00**	7,83	0,00**	22,52
m 14	0,00**	107,08	8,94	0,00**	8,16	0,00**	23,32
m 15	0,00**	107,70	8,09	0,00**	7,25	0,00**	19,36
m 16	0,00**	108,30	7,30	0,00**	5,98	0,00**	14,86
m 17	0,00**	108,49	6,82	0,00**	5,83	0,00**	13,55
m 18	0,00**	109,46	6,11	0,0**	4,00	0,00**	9,77
m 19	0,00**	109,87	5,30	0,00**	2,95	0,00**	6,74
m 20	0,00**	110,83	4,05	0,00**	1,31	0,16 ^{ns}	3,76
m 21	0,00**	110,29	3,36	0,00**	2,29	0,01*	3,02
m 22	0,00**	110,50	3,93	0,00**	1,98	0,03*	3,70
m 23	0,00**	110,43	4,64	0,00**	2,32	0,01*	5,22
m 24	0,00**	110,16	5,03	0,00**	3,02	0,00**	6,70
m 25	0,00**	109,98	5,50	0,00**	3,41	0,00**	8,10
m 26	0,00**	109,68	6,01	0,00**	3,95	0,00**	9,76
m 27	0,00**	109,12	6,94	0,00**	4,85	0,00**	13,00
m 28	0,00**	108,62	7,07	0,00**	5,63	0,00**	14,03
m 29	0,00**	108,34	7,08	0,00**	5,97	0,00**	14,26
m 30	0,00**	109,14	6,45	0,00**	4,28	0,00**	10,67
m 31	0,00**	109,20	6,02	0,00**	4,05	0,00**	9,17
m 32	0,00**	109,20	6,21	0,00**	4,07	0,00**	9,66
m 33	0,00**	109,56	5,96	0,00**	3,66	0,00**	8,93
m 34	0,00**	109,63	6,04	0,00**	3,33	0,00**	8,65
m 35	0,00**	110,25	6,26	0,00**	2,17	0,01*	8,59
m 36	0,00**	109,87	6,20	0,00**	2,93	0,00**	8,66
m 37	0,00**	109,95	5,24	0,00**	2,83	0,00**	6,46
m 38	0,00**	110,25	4,44	0,00**	2,24	0,01*	4,61
m 39	0,00**	110,31	4,22	0,00**	2,12	0,02*	4,16
m 40	0,00**	110,48	4,13	0,00**	1,70	0,07 ^{ns}	3,75
m 41	0,00**	111,00	3,68	0,00**	0,71	0,53 ^{ns}	2,72
m 42	0,00**	111,07	3,49	0,00**	0,53	0,58 ^{ns}	2,46
m 43	0,34 ^{ns}	111,72	0,95	0,16 ^{ns}	-0,52	0,59 ^{ns}	0,21
m 44	0,14 ^{ns}	111,23	1,40	0,04*	0,34	0,72 ^{ns}	0,38
m 45	0,09 ^{ns}	110,86	1,30	0,06 ^{ns}	1,08	0,26 ^{ns}	0,46
m 46	0,02*	110,62	1,60	0,02*	1,50	0,11 ^{ns}	0,74
m 47	0,00**	110,84	2,33	0,00**	1,10	0,25 ^{ns}	1,22
m 48	0,00**	110,06	2,93	0,00**	2,68	0,00**	2,56

Apêndice 3. Continuação.

m 49	0,00**	109,45	4,07	0,00**	3,77	0,00**	4,89
m 50	0,00**	109,11	3,83	0,00**	4,35	0,00**	4,93
m 51	0,00**	109,13	3,32	0,00**	4,49	0,00**	4,47
m 52	0,00**	109,01	2,88	0,00**	4,94	0,00**	4,51
m 53	0,00**	109,57	2,96	0,00**	3,84	0,00**	3,55
m 54	0,00**	109,46	3,22	0,00**	4,04	0,00**	4,07
m 55	0,00**	109,61	3,50	0,00**	3,63	0,00**	4,02
m 56	0,00**	109,47	3,69	0,00**	3,84	0,00**	4,43
m 57	0,00**	109,34	3,45	0,00**	4,09	0,00**	4,33
m 58	0,00**	109,89	3,29	0,00**	3,05	0,00**	3,27
m 59	0,00**	110,48	2,37	0,00**	1,83	0,05 ^{ns}	1,50
m 60	0,04*	110,85	1,52	0,03*	1,16	0,23 ^{ns}	0,61
m 61	0,03*	110,68	1,42	0,04*	1,49	0,12 ^{ns}	0,67
m 62	0,12 ^{ns}	111,04	1,24	0,07 ^{ns}	0,85	0,62 ^{ns}	0,41
m 63	0,11 ^{ns}	111,26	1,43	0,03*	0,40	0,68 ^{ns}	0,44
m 64	100,00 ^{ns}	111,34	0,09	0,88 ^{ns}	0,20	0,82 ^{ns}	0,00
m 65	100,00 ^{ns}	111,50	0,32	0,64 ^{ns}	-0,10	0,91 ^{ns}	0,02
m 66	100,00 ^{ns}	111,75	0,51	0,54 ^{ns}	-0,58	0,55 ^{ns}	0,09
m 67	100,00 ^{ns}	111,14	-0,14	0,83 ^{ns}	0,60	0,54 ^{ns}	0,04
m 68	100,00 ^{ns}	111,41	-0,23	0,73 ^{ns}	0,05	0,95 ^{ns}	0,01
m 69	100,00 ^{ns}	111,71	0,47	0,50 ^{ns}	-0,50	0,61 ^{ns}	0,07
m 70	100,00 ^{ns}	111,28	0,57	0,59 ^{ns}	0,37	0,70 ^{ns}	0,09
m 71	100,00 ^{ns}	111,24	0,41	0,54 ^{ns}	0,44	0,65 ^{ns}	0,06
m 72	100,00 ^{ns}	111,80	-0,04	0,94 ^{ns}	-0,69	0,52 ^{ns}	0,05
m 73	100,00 ^{ns}	111,73	-0,39	0,57 ^{ns}	-0,56	0,56 ^{ns}	0,06
m 74	100,00 ^{ns}	111,99	-0,47	0,50 ^{ns}	-1,06	0,26 ^{ns}	0,16
m 75	100,00 ^{ns}	111,60	-0,12	0,85 ^{ns}	-0,30	0,75 ^{ns}	0,01
m 76	100,00 ^{ns}	111,21	-0,30	0,66 ^{ns}	0,45	0,64 ^{ns}	0,04
m 77	100,00 ^{ns}	111,09	-0,36	0,61 ^{ns}	0,67	0,51 ^{ns}	0,07
m 78	100,00 ^{ns}	111,58	-0,22	0,74 ^{ns}	-0,27	0,77 ^{ns}	0,01
m 79	100,00 ^{ns}	111,26	0,17	0,79 ^{ns}	0,36	0,71 ^{ns}	0,02
m 80	100,00 ^{ns}	111,37	-0,19	0,78 ^{ns}	0,14	0,87 ^{ns}	0,01
m 81	100,00 ^{ns}	111,49	-0,08	0,90 ^{ns}	-0,09	0,92 ^{ns}	0,00
m 82	100,00 ^{ns}	111,66	0,28	0,68 ^{ns}	-0,40	0,68 ^{ns}	0,03
m 83	100,00 ^{ns}	111,29	0,45	0,51 ^{ns}	0,32	0,74 ^{ns}	0,05
m 84	0,25 ^{ns}	111,86	0,91	0,17 ^{ns}	-0,88	0,63 ^{ns}	0,27

^{ns}, * e ** - Não significativos, significativos a 5% e 1%, respectivamente.