

EDUARDO MACHADO CRUZ

EFEITO DA SATURAÇÃO E DO TAMANHO DE POPULAÇÕES F₂ E DE
RETROCRUZAMENTO SOBRE A ACURÁCIA DO MAPEAMENTO
GENÉTICO

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

VIÇOSA
MINAS GERAIS – BRASIL
2006

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

T

C957e
2006

Cruz, Eduardo Machado, 1975.

Efeito da saturação e do tamanho de populações
F₂ e de retrocruzamento sobre a acuária do
mapeamento genético / Eduardo Machado Cruz.
– Viçosa : UFV, 2006.
x, 119f. : il. ; 29cm.

Orientador: Cosme Damião Cruz.

Tese (doutorado) - Universidade Federal de
Viçosa.

Referências bibliográficas: f. 114-119.

1. Mapeamento cromossômico. 2. Genoma - Efeito
da saturação. 3. Genoma - Simulação por computador.
4. GENES (Programa de computador). 5. GQMOL
(Programa de computador). I. Universidade Federal de
Viçosa. II. Título.

CDD 22.ed. 572.8633

EDUARDO MACHADO CRUZ

EFEITO DA SATURAÇÃO E DO TAMANHO DE POPULAÇÕES F_2 E DE
RETROCRUZAMENTO SOBRE A ACURÁCIA DO MAPEAMENTO
GENÉTICO

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

APROVADA: 20 de abril de 2006.

Prof. Paulo Roberto Cecon
(Conselheiro)

Prof. Pedro Crescêncio Souza Carneiro
(Conselheiro)

Prof. Eduardo Rezende Galvão

Prof. Sebastião Martins Filho

Prof. Cosme Damião Cruz
(Orientador)

AGRADECIMENTO

A Deus.

À minha família, por estar sempre comigo.

Ao Professor Cosme Damião Cruz, pela enorme paciência, pela amizade, pela confiança, pelo incentivo e pela orientação.

À Universidade Federal de Viçosa (UFV), pela oportunidade de realização deste curso.

À Coordenação de Aperfeiçoamento do Pessoal de Nível Superior (CAPES) e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pelas bolsas de estudo concedidas.

Aos professores de graduação e de pós-graduação, pela atenção, pela disponibilidade e pelos ensinamentos.

Aos meus grandes amigos Ana Paula, João Francisco, Samuel e Vander, pela amizade, pelo apoio e pelo companheirismo.

Aos meus demais amigos e colegas da UFV, pela convivência.

A todos que, direta ou indiretamente, contribuíram para a realização deste trabalho.

Finalmente, a quem estiver valorizando este trabalho através de sua leitura e utilização de alguma forma.

BIOGRAFIA

EDUARDO MACHADO CRUZ, filho de José Carlos Cruz e Elva do Carmo Machado Cruz, nasceu em Belo Horizonte, Minas Gerais, em 25 de março de 1975.

Em 13 de agosto de 1999, graduou-se em Agronomia pela Universidade Federal de Viçosa (UFV), em Viçosa, Minas Gerais, e, em seguida, ingressou no Programa de Pós-Graduação, em nível de Mestrado, em Genética e Melhoramento dessa na mesma Instituição, obtendo, em 14 de dezembro de 2001, o título de Magister Scientiae.

Em 2002, iniciou-se no Programa de Pós-Graduação, em nível de Doutorado, em Genética e Melhoramento da UFRV, submetendo-se à defesa de tese em abril de 2006.

CONTEÚDO

	Página
RESUMO.....	vii
ABSTRACT.....	ix
1. INTRODUÇÃO.....	1
2. REVISÃO DE LITERATURA.....	3
2.1. Análise genômica.....	3
2.2. Marcadores.....	4
2.3. Importância dos marcadores moleculares no melhoramento genético	7
2.4. Mapeamento.....	11
2.4.1. Tamanho da população ou da amostra.....	11
2.4.2. Número de marcadores	12
2.5. Simulação.....	18
2.6. Contribuição dos estudos de simulação na análise genômica	23
2.7. Programas GENES e GQMOL.....	25
3. MATERIAL E MÉTODOS.....	27
3.1. Simulação de dados	27
3.1.1. Simulação do genoma.....	27

	Página
3.1.2 Simulação dos genitores	28
3.1.3. Tamanho da população	30
3.1.4. Procedimento de simulação dos indivíduos das populações F ₂ e de retrocruzamento	30
3.2. Análise genômica – Mapeamento.....	31
3.2.1. Análise de segregação de locos individuais	31
3.2.2. Estimação da porcentagem de recombinação	31
3.2.2.1. População de retrocruzamento.....	32
3.2.2.2. População F ₂ com marcadores co-dominantes	33
3.3. Comparação de genomas	36
3.3.1. Número de grupos de ligação e marcas por grupo	37
3.3.2. Tamanho do grupo de ligação.....	37
3.3.3. Mé...dia das distâncias entre dois marcadores adjacentes no grupo de ligação.....	37
3.3.4. Variância das distâncias entre marcas adjacentes.....	38
3.3.5. Correlação de Spearman	38
3.3.6. Estresse	40
3.4. Testes de comparação múltipla.....	41
4. RESULTADOS E DISCUSSÃO.....	42
4.1. Considerações gerais.....	42
4.2. Índice de informatividade e desvio-padrão em populações F ₂	50
4.3. Recuperação de grupos de ligação em populações F ₂	51
4.3.1. Saturação dos grupos de ligação.....	51
4.3.2. Correlação de Spearman entre medidas de distância.....	58
4.3.3. Tamanho dos grupos de ligação	60
4.3.4. Média das distâncias entre marcas adjacentes.....	64
4.3.5. Variância das distâncias entre marcas adjacentes.....	69
4.3.6. Estresse	74
4.4. Índice de informatividade e desvio-padrão em populações de retrocruzamento	79

	Página
4.5. Recuperação de grupos de ligação em populações de retrocruzamento.....	81
4.5.1. Saturação de grupo de ligação	82
4.5.2. Correlação de Spearman entre medidas de distância.....	89
4.5.3. Tamanho dos grupos de ligação	92
4.5.4. Média das distâncias entre marcas adjacentes.....	96
4.5.5. Variância das distâncias entre marcas adjacentes.....	101
4.5.6. Estresse	106
5. CONCLUSÕES	112
REFERÊNCIAS.....	114

RESUMO

CRUZ, Eduardo Machado, D. S., Universidade Federal de Viçosa, abril de 2006.

Efeito da saturação e do tamanho de populações F_2 e de retrocruzamento sobre a acurácia do mapeamento genético. Orientador: Cosme Damião Cruz. Conselheiros: Paulo Roberto Cecon, Pedro Crescêncio Souza Carneiro e Luiz Alexandre Peternelli.

Este trabalho objetivou gerar e analisar dados de genomas simulados de populações F_2 e de retrocruzamento e, com base nesses dados, avaliar o tamanho ótimo de amostras dessas populações para estudo de mapeamento genético. Para isso, foram geradas populações F_2 e de retrocruzamento com sete tamanhos de amostra (50, 100, 150, 200, 400, 600 e 800 indivíduos) e quatro níveis de saturação (5, 10, 15 e 20 marcadores em cada um de 10 grupos de ligações de 100 centimorgans). As simulações foram feitas por meio do programa GQMOL (2005) e analisadas com os programas GENES (2004) e GQMOL (2005), criados pelo professor Cosme Damião Cruz, do Departamento de Biologia da Universidade Federal de Viçosa. Conclui-se que é completamente inviável a utilização de amostras pequenas, com tamanho próximo a 50 indivíduos, mesmo na saturação de 5 cM por grupo de ligação, que foi o maior nível de saturação utilizado neste trabalho, e desnecessário o emprego de amostras a partir de 400 indivíduos para obtenção de mapas genéticos acurados. Amostras com pequeno

número de indivíduos não permitem recuperar o número correto de grupos de ligações da espécie e o ordenamento correto, nem possibilitam estimar corretamente a distância entre os pares de locos. Em populações F_2 , tamanhos mínimos de amostras de 100, 150 e 200 indivíduos podem ser utilizados para uma recuperação completa do genoma com saturações de 5, 6,66 e 10 cM, respectivamente. No caso de populações de retrocruzamento de 200 indivíduos para os três primeiros níveis de saturação e de 800 para a saturação de 20 cM e de existirem poucos indivíduos, porém acima de 100, para que seja feita uma avaliação, mapas confiáveis podem ser gerados desde que se utiliza uma saturação superior a 5 cM.

ABSTRACT

CRUZ, Eduardo Machado, D. S., Universidade Federal de Viçosa, April, 2006.
Effect of saturation and size of F2 and backcrossed populations on genetic mapping accuracy. Adviser: Cosme Damião Cruz. Committee Members: Paulo Roberto Cecon, Pedro Crescêncio Souza Carneiro and Luiz Alexandre Peternelli.

The objective of this work was to generate and analyze data of simulated genomes of F2 and backcrossed populations, and based on these data, to evaluate the optimum size of population samples for genetic mapping studies. F2 and backcrossed populations with seven sample sizes (50, 100, 150, 200, 400, 600 and 800 individuals) and four levels of saturation were generated (5, 10, 15 and 20 markers in each of the 10 linkage groups of 100 centimorgans). Simulations were performed using the GQMOL software (2005) and analyzed with both GENES (2004) and GQMOL (2005) software developed by Professor Cosme Damião Cruz, from the Department of General Biology at the Federal University of Viçosa, Viçosa, MG. The results led to the conclusion that the use of small samples, approximately 50 individuals, is completely impracticable, even with saturation of 5 cM per linkage group, which was the highest level of saturation used in this work, as well as it is unnecessary the use of samples above 400

individuals to attain map accuracy. Samples with small number of individuals do not allow the recovering of the exact number of species linkage groups and the correct order, neither the accurate estimation of the distance between pairs of loci. For F2 populations, the smallest sample sizes of 100, 150 and 200 individuals can be used for a complete recovery of the genome with saturations of 5; 6.66; and 10 cM respectively, and for backcrossed populations of 200 individuals in the first three levels of saturation and 800 in the saturation of 20 cM. If few individuals exist, above 100 though, to carry out an evaluation, reliable maps can be constructed, since saturation above 5 cM is used.

1. INTRODUÇÃO

O mapeamento genético facilita o trabalho de melhoramento, uma vez que uma ou mais marcas do genótipo podem ser associadas a um ou mais genes controladores de características, qualitativas e quantitativas (QTL). Desse modo, tendo-se o genótipo mapeado, o trabalho de melhoramento poderá ser otimizado, tanto na eficiência do programa quanto na velocidade de obtenção de ganhos. Entretanto, a disponibilidade de um mapa genético fidedigno depende de uma série de fatores como tipo de marcador utilizado, tipo de população analisada e o tamanho da população, dentre outros. Além disso, alguns aspectos de natureza metodológica também devem ser considerados no processo de obtenção do mapa genético, como análise de segregação de locos individuais, análise de segregação conjunta, níveis mínimos de recombinação e máximos de LOD scores (para estabelecimento de ligação e rotinas de ordenamento de locos, dentre outras). Com isso, nota-se que uma das ferramentas disponíveis, e que vem se tornando cada vez mais necessária, ao melhorista é a Genômica.

A Genômica pode ser entendida como uma denominação dada à ciência que estuda o genoma de forma completa, integrando diversas áreas da Genética, como a Genética mendeliana, a Citogenética, a Genética Molecular, a Genética de Populações e a Genética Quantitativa; e outras áreas do conhecimento, como a Ciência da Computação e os Sistemas Automatizados.

Através da incorporação da Ciência da Computação e de Sistemas Automatizados ao processo de melhoramento, passou a ser possível a análise e comparação de técnicas de mapeamento, estudo da seleção assistida e geração de amostras de comportamento semelhante a diferentes tipos de população, assim como simular um genoma inteiro, além do aumento da eficiência dos programas de melhoramento. Essas e outras vantagens da simulação eliminam gastos, que passaram a ser desnecessários, ao longo do processo de mapeamento.

Os objetivos deste trabalho foram gerar e analisar dados a partir de um programa de simulação de genoma e de populações e, com base nesses dados simulados, avaliar o tamanho ótimo de populações F_2 e de retrocruzamento, para estudo de mapeamento genético.

2. REVISÃO DE LITERATURA

2.1. Análise genômica

Com o avanço do melhoramento genético, a ciência passou a desempenhar papel fundamental no desenvolvimento de plantas e animais de importância para o homem. Autores como Borém (1997) e Nass et al. (2001) descreveram áreas do conhecimento, nos diversos ramos da Genética, inclusive os marcadores genéticos de DNA, que são de extrema importância para o melhoramento e possibilitam torná-lo cada vez mais eficiente na busca da melhoria de caracteres de interesse.

Requisitos de produtividade, uniformidade para permitir o uso de máquinas para colheita e beneficiamento, adaptação a novos ambientes para permitir a expansão de fronteiras agrícolas, resistência a doenças e pragas foram, e deverão ser, os caracteres de maior importância a serem considerados por melhoristas e geneticistas. Em tempos mais recentes, contudo, os consumidores também passaram a influenciar o direcionamento das ações do melhoramento. Estes últimos anseiam por alimentos com maior qualidade nutricional, sabores melhores e diferentes, boa aparência, frutos sem sementes e em tamanhos diferentes, animais com menores teores de gordura e preços competitivos, o que novamente tem íntimo relacionamento com os aspectos de produtividade e outros

itens relevantes na redução de custos de produção. Todas essas características estão passando a ter papel decisivo na aceitação de um produto, obrigando o melhorista a estabelecer estratégia de visão de futuro sobre seu trabalho, principalmente enfocando tanto aspectos do negócio agrícola e suas implicações para o melhoramento quanto a qualidade nutricional e de aceitação dos produtos. Por isso, uma das ferramentas disponíveis, e que vem se tornando cada vez mais necessária ao melhorista e cada vez mais utilizada, é a Genômica (CRUZ *apud* NASS, 2001).

Ao trabalhar com a Genômica, atua-se em pelo menos três áreas: na Genômica Clássica, em que são estudados os marcadores genéticos, quando se fazem análises de ligação, o ordenamento de genes nas análises multipontos, o mapeamento genético e o de QTLs e a avaliação de seus efeitos; na Informática, que investe na formação de banco de dados, em que são realizadas comparações de seqüências e são promovidas comunicações de dados e automação de maneira geral; e nas Análises de Seqüências de DNA, em que são realizados seqüenciamentos, alinhamentos e comparações de seqüências obtidas na Genômica Clássica, além de estudos de simulação para se identificar a viabilidade de determinado modelo estatístico a ser aplicado em programas de mapeamento (SCHUSTER e CRUZ, 2004).

2.2. Marcadores

Por volta dos anos de 1960, os marcadores utilizados em estudos de genética e melhoramento eram aqueles determinados por locos associados a características morfológicas, sendo que geralmente essas características eram de fácil identificação visual. Essa estratégia de trabalho utilizando marcadores fenotípicos foi proposta por Sax (1923). Porém, além de alguns marcadores morfológicos, causam efeitos indesejáveis, e os marcadores moleculares apresentam vantagens como alto nível de polimorfismo para cada loco estudado. Isso facilita o desenvolvimento de mapas a partir de populações segregantes e de cruzamentos específicos, a neutralidade em relação aos efeitos de ambiente, com

pouco ou nenhum efeito de epistasia ou pleiotropismo, e são em geral co-dominantes, permitindo a obtenção de maior quantidade de informação sobre o loco estudado (FERREIRA e GRATAPAGLIA, 1996).

A revolução no campo de marcadores iniciou-se com o desenvolvimento dos marcadores isoenzimáticos. A partir daí, o número de marcadores genéticos foi grandemente ampliado, e a aplicabilidade da técnica passou a ser geral (TANKSLEY et al., 1989). Contudo, o número de marcadores previsto pelos ensaios isoenzimáticos é limitado, ficando geralmente entre 10 e 20 por espécie. Mesmo considerando que o número total de locos isoenzimáticos que podem ser detectados seja maior que 100 (MURPHY, 1990 *apud* FERREIRA e GRATAPAGLIA, 1996), o nível de resolução dos marcadores isoenzimáticos não permitia cobertura completa do genoma, limitando certos estudos, como a construção de mapas genéticos.

Os marcadores de DNA possuem algumas vantagens em relação aos marcadores morfológicos e isoenzimáticos. Uma primeira vantagem é que são baseados na variação de seqüências de nucleotídeos, sendo assim numericamente ilimitados; quando co-dominantes, permitem a identificação de genótipos em qualquer tipo de cruzamento. Uma segunda vantagem é que são fenotipicamente neutros e não possuem interação epistática, quando ocorre interação entre diferentes genes, não-alélicos, sobre uma mesma característica. A utilidade desses marcadores está em estabelecer relações de ligação entre eles e os locos de interesse. Essa ligação permite, ao pesquisador, inferir sobre a presença dos locos, mediante o ensaio de alguns desses marcadores, facilitando trabalhos de transferência e de mapeamento (TANKSLEY et al., 1989).

Estudos fundamentados apenas na genética quantitativa clássica não permitem identificação direta de genes ou locos de caracteres quantitativos (QTLs), mas suas posições podem ser estimadas com a ajuda de marcadores genéticos. Pesquisadores, na área de Biologia Molecular, podem avaliar se um gene, supostamente ligado a uma expressão, esteja realmente ligado, através da análise de QTLs, desde que este corresponda à expressão de um particular gene já mapeado e cuja expressão primária seja conhecida (LEBRETON et al., 1998).

Segundo Soller e Beckmann (1983), quando marcadores co-dominantes (isoenzimáticos ou RFLPs) estão disponíveis, análises baseadas em gerações F_2 serão mais úteis que aquelas baseadas em gerações de retrocruzamento, por fornecerem maiores informações tanto em relação à dominância quanto a um efeito maior do QTL identificado.

Estudos de autores como Weller (1986), Patterson et al. (1988) e outros têm indicado que grande variedade de efeitos quantitativos pode ser associada a marcadores genéticos específicos. O método de máxima verossimilhança e outros métodos de mapeamento atestam que os resultados são, em maior parte, devidos à ligação e não à pleiotropia. Esses experimentos indicam que uma porção significativa dos efeitos associados a marcadores em características de natureza quantitativas são de uma magnitude que poderia ser prontamente detectada em experimentos com aproximadamente 1.000 indivíduos de população F_2 ou de retrocruzamento. É importante notar que se duas linhagens diferem em muitos marcadores, localizados ao longo do genoma, uma única geração F_2 permitirá que se façam uma avaliação total e o mapeamento do genoma inteiro de todas as características quantitativas avaliadas. A quantidade de informação produzida por esse tipo de análise é crítica, dependendo do número de marcadores polimórficos que diferem entre as linhas parentais.

Em relação à aplicabilidade de marcadores em programas de melhoramento genético, dois grupos de espécies autofecundantes são distinguidos. O primeiro grupo é composto por espécies autógamas de genoma pequeno (~10 Morgans). Nesse grupo, análises auxiliadas por marcadores genéticos podem ser feitas, usando-se delineamentos eficientes, e relativamente pouco polimorfismo é necessário para que o genoma inteiro seja coberto. Desse modo, os marcadores genéticos podem contribuir de modo decisivo para o melhoramento dessas espécies. O segundo grupo é composto de populações de cruzamento aberto, incluindo animais, de genoma grande (~30 Morgans). Nessas espécies, métodos de seleção assistida por marcadores não são de aplicabilidade geral, mas podem ter uma contribuição significativa em situações particulares (SOLLER e BECKMANN, 1983).

A aplicação da técnica de marcadores genéticos às populações segregantes com genoma de grande tamanho é problemática. Apesar disso, duas aplicações podem ser antevistas. A primeira é a seleção precoce de machos, no caso de gado leiteiro, por exemplo, de acordo com a avaliação cromossômica de seus pais. Avaliação de custos baseados em RFLP indica que essa aplicabilidade pode ser economicamente justificável (BECKMANN e SOLLER, 1983). Uma segunda aplicação potencial envolve a análise genética do cruzamento entre uma linhagem caracterizada por alto valor quantitativo para tratamentos de baixo valor de herdabilidade, porém de grande importância econômica, e uma linhagem comercial padrão. Nesses casos é esperada uma grande diferença na frequência do QTL entre a linha comercial e a fonte do gene de interesse, possibilitando que seja feita uma análise relativamente barata (SOLLER e BECKMANN, 1983).

Quando um único tratamento é de interesse, pode-se conseguir uma considerável redução no número de indivíduos escolhidos para marcadores polimórficos, através da seleção de marcadores genéticos apenas em indivíduos que apresentarem alta ou baixa expressão da característica quantitativa. Se a linhagem-fonte estiver com marcadores identificados associados ao QTL de interesse, essa linhagem poderá ser usada para introgridir o QTL em linhas comerciais (SOLLER e BECKMANN, 1983).

2.3. Importância dos marcadores moleculares no melhoramento genético

A seleção e recombinação de genótipos superiores identificados com o auxílio de marcadores moleculares têm sido defendidas e cada vez mais recomendadas pela literatura desde o surgimento de técnicas moleculares, por possibilitarem um estudo genético mais detalhado (FERREIRA e GRATTAPAGLIA, 1998). Essa prática tem-se tornado cada vez mais frequente com relação a características de herança simples, entretanto, para características quantitativas, algumas dificuldades ainda estão sendo enfrentadas. Como as características quantitativas são controladas por grande número de genes de pequeno efeito, que são influenciados pelo ambiente, a identificação dos locos

das características quantitativas é mais complexa que a identificação de locos para características qualitativas, ou de herança simples.

Marcadores ligados próximos a genes de interesse podem ser identificados pela técnica de análise de “bulk” segregante (BAS) (MICHELMORE et al., 1991).

Hospital et al. (1992) estudaram os efeitos de tempo, intensidade de seleção, tamanho da população, número e posição dos marcadores selecionados para cromossomos carregando, ou não, gene introgressido. Desse modo, avaliaram o uso de marcadores para acelerar a recuperação de genótipos recorrentes durante programas de introgressão. Ainda segundo esses autores, marcadores são mais úteis quando sua posição no mapa é conhecida.

Visscher et al. (1996), utilizando técnicas de simulação, avaliaram a eficiência da introgressão de genes em populações de retrocruzamento derivados de linhas endogâmicas. “Background genotypes” foram simulados assumindo um modelo genético de vários genes de pequeno efeito em fase de aproximação, que explicam a diferença entre os descendentes observados e a variância nas populações de retrocruzamento. Após a obtenção dos dados, constatou-se que marcadores foram eficientes tanto para introgressão de alelos quanto para seleção de genótipos receptores, simultaneamente. Ao serem consideradas marcas espaçadas de 10 a 20 cM, houve a vantagem de uma a duas gerações de seleção de retrocruzamento, em relação ao tempo gasto no melhoramento clássico, em relação à seleção fenotípica. Até mesmo no caso de QTLs mapeados de forma relativamente precisa, marcadores flanqueadores ou marcadores haplótipos devem cobrir de 10 a 20 cM próximos à posição estimada do gene, para que seja assegurado que a frequência do alelo não decline nas futuras gerações de retrocruzamento. Os resultados foram obtidos considerando-se cruzamentos entre linhas endogâmicas com marcadores 100% informativos, e em todos os cálculos foi usada a função de mapeamento de Haldane (1919), que admite interferência nula. Para modelagem do genótipo, em diferentes gerações de retrocruzamento foi escolhido um modelo infinitesimal de locos ligados para “background genotype”, em que grande número de locos ligados, todos de efeito pequeno e

igual e em fase de aproximação em linhas endogâmicas, explicasse as diferenças entre os descendentes e da variância no primeiro retrocruzamento ou numa população F_2 .

Brahm et al. (2000), trabalhando com o desenvolvimento de marcadores moleculares para o gene Pl2 em cultivares de girassol, visando à resistência a *Plasmopara hasteei*, raças 1, 2, 7 e 9, utilizaram dois conjuntos de linhagens isogênicas muito próximas (AS110/AS110Pl2 e S1358/S1358 Pl2) e um “bulk” de populações F_2 segregantes, para identificação de Random Amplified Polimorphic DNA (RAPD) e Amplified Fragment Length Polymorphisms (AFLP). Foi observado que os marcadores RAPD OPAA14750 e OPAC20831, assim como os marcadores AFLP E35M48-3, mostraram estar ligados bem próximos ao gene Pl2, aproximadamente a 2 cM. É comentado pelos referidos autores que os resultados apontam a aplicabilidade na seleção assistida por marcadores, permitindo a diferenciação entre germoplasmas resistentes e suscetíveis em um conjunto de diversas linhagens endogâmicas. Ainda é comentado por esses autores que a idéia de se introduzir gene resistente a todas as raças de um patógeno pode ser muito atraente, porém não é recomendável. Em vez disso, devem-se utilizar cultivares híbridos com resistência a diferentes raças do patógeno, mantendo baixa a pressão de seleção e aumentando, desse modo, o ciclo de cada gene de resistência.

Em se tratando de espécies arbóreas, além dos problemas normalmente enfrentados no melhoramento, fatores como o tempo, a escassez de informações genéticas e a alogamia dificultam o trabalho dos pesquisadores. É visto que a seleção assistida por marcadores diminui esses problemas, tendo potencial para aumentar a eficiência dos programas de melhoramento, desde que as características de interesse sejam controladas por um único gene dominante (KONDO et al., 2000).

Segundo visto no trabalho de Frish et al. (1999), a seleção assistida por marcadores pode acelerar a recuperação do genótipo do parental recorrente, em programas de melhoramento, com retrocruzamento. Nesse estudo, em que foram feitas simulações computacionais para comparar estratégias para seleção assistida

por marcadores para retrocruzamento de um gene, para proporção de Recurrente Parent Genome (RPG) recuperado e para o número de MDP (marker data points) necessários para um programa elaborado para introgressão de um alelo desejado em uma linhagem. As simulações foram feitas com base em um mapa genético de milho considerando 80 marcas, e a seleção para o alelo desejado foi baseada na avaliação fenotípica. Em comparação com uma população de tamanho constante, uma população de tamanho crescente de retrocruzamento 1 para retrocruzamento 2 diminuiu o número de MDP, por volta de 50%, sem afetar a proporção de RPG. Um método de seleção de quatro estágios, enfatizando a seleção para recombinantes no cromossomo portador do alelo de interesse nas primeiras gerações, reduziu o número de MDP necessários, cerca de 75%, em comparação com o método de seleção de referência. Adotados os princípios citados, para o desenho de um programa de retrocruzamento assistido por marcadores houve uma economia substancial em números de MDP necessários.

Schuster et al. (2001) estudaram QTLs associados à resistência ao nematóide do cisto da soja (*Heterodera glycines*), encontrando quatro marcadores RAPD e quatro marcadores microssatélites. Os marcadores RAPD foram responsáveis por mais de 40% da resistência ao nematóide da raça 14. No caso dos microssatélites, foram identificados genótipos homozigotos resistentes, sendo a eficiência ainda maior.

Marcadores baseados na reação de polimerase em cadeia, como RAPD e microssatélites, estão sendo cada vez mais usados por serem mais abundantes, menos dispendiosos, mais polimórficos e menos trabalhosos para serem manipulados que RFLP (SCHUSTER et al., 2001).

Trabalhando com seleção assistida por marcadores para resistência ao míldio pulverulento (*Uncinula necator* (Shaw.) Burr) em videiras, Dalbó et al. (2001) construíram um mapa genético através da utilização de uma população segregante originada de “Horizon” x Illinois 547-1, sendo que esse cruzamento representa a formação de um híbrido entre *Vitis rupestris* Scheele e *Vitis cinérea* Engelm. O QTL obtido a partir do parental resistente, Illinois 547-1, explicou 41% das variações observadas. A taxa de distorção da segregação foi observada

em alguns cruzamentos, casos em que o alelo estava associado ao QTL de resistência. Ao ser feita a seleção com base nos marcadores, a taxa de indivíduos suscetíveis selecionados diminuiu. É também comentado que a eficiência da seleção, nos cruzamentos, foi maior onde a distorção foi mais intensa.

2.4. Mapeamento

Primeiramente na construção de um mapa genético, deve ser feito o estudo individual dos locos que devem apresentar segregação mendeliana em função do tipo de população a ser estudada. Nas populações de retrocruzamento, duplo-haplóides e RILs, espera-se uma razão de segregação de 1:1 e nas populações F_2 , uma razão de segregação de 1:2:1, para locos co-dominantes, ou 3:1, para locos dominantes. Para verificar se a segregação de um loco ocorre como esperado, é feita uma comparação do número de indivíduos observados em cada classe com o esperado de acordo com a razão de segregação. Um teste que se tem mostrado bastante eficiente e útil para esse fim é o do qui-quadrado (χ^2). Além de levar em consideração os desvios ocorridos entre os valores observados e esperados, esse teste também é sensível ao tamanho da amostra.

2.4.1. Tamanho da população ou da amostra

Um dos fatores de fundamental importância para se obterem dados consistentes em um trabalho de mapeamento é o tamanho da amostra ou população a ser trabalhada. A resolução do mapa e a capacidade de se determinar a seqüência de marcadores no mapa trabalhado é diretamente relacionada ao tamanho da amostra ou população. É certo que as amostras com menos de 50 indivíduos provavelmente terão baixa resolução de mapeamento, principalmente na detecção de QTLs de pequeno efeito (YOUNG, 1994).

Populações com poucos genótipos podem não permitir a observação da quebra de ligação entre os marcadores e a conseqüente determinação da distância entre estes. O tamanho adequado da população também depende do tipo de população, sendo que populações de retrocruzamento contêm aproximadamente a

metade do conteúdo informativo de populações F_2 . A grande vantagem do uso de populações F_2 é que os genótipos dessas populações são formados por dois gametas que têm seus locos em desequilíbrio de ligação, e não por um único, como no caso dos genótipos obtidos por retrocruzamento (LANNES, 2004).

Lannes et al. (2004), em estudo com arroz, composto por duas populações de retrocruzamento, uma com 53 e outra com 74 indivíduos, utilizando 138 “primers”, sendo 131 RAPD obtidos da University of British Columbia e sete que haviam amplificado bandas polimórficas mais consistentes entre os pais, comentaram que os mapas genéticos obtidos poderiam ser considerados como básicos, uma vez que o número de marcadores mapeados e os tamanhos das populações de mapeamento foram pequenos. Outra observação feita por esses autores foi a presença de sete intervalos que apresentaram distância de 0 cM entre marcadores adjacentes. Segundo eles, isso pode ter acontecido devido ao pequeno tamanho da população (53 indivíduos) utilizada para a formação desse mapa.

2.4.2. Número de marcadores

Sabe-se que o desenvolvimento da metodologia de construção de mapas genéticos remonta do início do século passado. Após a redescoberta do trabalho de Mendel em 1900, 34 anos depois de sua publicação várias pesquisas foram realizadas com o objetivo de ampliar e validar suas conclusões em relação ao mecanismo de herança de características quantitativas. Bateson e Punnet (1905), citados por Stutervant (1965), trabalhando com características cor da flor e formato do grão de pólen, em ervilha, publicaram um dos primeiros relatos da ligação gênica (ROCHA, 2003), entretanto a primeira evidência de que os genes estão localizados em posições definidas foi relatada por Morgan (1910), em trabalho com análise do padrão de herança de um gene mutante ligado ao sexo, em *Drosophila melanogaster*.

O mapeamento genético baseado na análise da frequência dos genótipos foi idealizado por Stutervant (1913), que publicou o primeiro mapa genético.

Alfred H. Stutervant, trabalhando com seis genes ligados ao sexo em *Drosophila melanogaster*, não apenas produziu o primeiro mapa genético com todos os genes na sua posição correta, como também propôs o princípio básico do mapeamento genético e da utilização da frequência de recombinantes para estimar a distância entre dois genes. Em homenagem a seu orientador Thomas H. Morgan, determinou a unidade de mapa de centiMorgan (cM).

É fundamental que se calcule a distância entre genes, estabelecendo um ordenamento e formando grupos de ligações. Como os eventos de permutação ocorrem ao acaso ao longo do cromossomo, a probabilidade de recombinação é maior em locos que se encontram a uma maior distância entre si do que naqueles mais próximos. Isso pode ser considerado a idéia básica do mapeamento genético, ou seja, a taxa de recombinação entre os locos é usada como referência para ordenamento dos genes (ou marcadores) nos cromossomos (SCHUSTER e CRUZ, 2004).

De modo prático, a frequência de recombinação entre dois locos é obtida pelo número de recombinantes dividido pelo total de indivíduos analisados. Assim, conhecendo-se as frequências de recombinação entre diversos locos do mesmo grupo de ligação, é possível estimar a ordem dos locos nesse grupo. Ressalta-se que um grupo de ligação é um conjunto de marcadores genéticos que possuem menos de 50% de recombinação entre dois marcadores consecutivos (SCHUSTER e CRUZ, 2004).

A falta de aditividade das frequências de recombinação levou ao desenvolvimento de funções de mapeamento, as quais são utilizadas para converter frequências de recombinação em medidas de distâncias com propriedades mais interessantes para o ordenamento de locos. Das funções de mapeamento, as mais conhecidas são as de Haldane (1919), que admitiu a independência das permutas nos intervalos adjacentes, e a de Kosambi (1944), que considerou a interferência.

A função de Haldane é expressa por:

$$m = \frac{-\ln(1-2r)}{2}$$

em que m é a distância em Morgan. Para obter a distância em centiMorgan (cM), basta multiplicar por 100.

Para obter a frequência de recombinação a partir da distância em Morgan, utiliza-se a expressão:

$$r = \frac{1 - e^{-2m}}{2}$$

O mapa de ligação de uma espécie pode ser definido como um arranjo linear de um grupo de genes ou marcadores adjacentes. Dois marcadores são ditos ligados sempre que menos de 50% dos gametas produzidos apresentam genótipos recombinantes para esses dois genes (STUTERVANT, 1913).

Para Ooijen (1992), o desenvolvimento de mapas de ligação com grande número de marcadores moleculares tem estimulado a busca por métodos para mapeamento de genes envolvidos em características quantitativas (QTL). Um método promissor proposto por Lander e Botstein (1989), já citado, emprega pares de marcadores vizinhos para obter o máximo de informação da ligação de QTLs dentro do segmento de cromossomo analisado.

Neste trabalho, a acurácia desse método foi investigada por simulação computacional. Os resultados aqui obtidos indicam que existe uma probabilidade razoável de detecção de QTL que explique pelo menos 5% da variância. Tanto o número de indivíduos quanto o tamanho relativo do efeito genotípico do QTL são fatores importantes na determinação da precisão do mapeamento. Na média, um QTL com capacidade para explicar 5 ou 10% da variância é mapeado com um intervalo de 40 ou 20 cM, respectivamente. É claro que QTLs com maiores efeitos genotípicos serão localizados mais precisamente. Contudo, deve ser notado que o comprimento do intervalo é variável.

Devido ao grande número de fatores que influenciam a acurácia do mapeamento de QTL, este estudo se restringiu a alguns poucos casos simples. Uma primeira geração de retrocruzamento e uma população F_2 foram estudadas, sendo os parentais homozigotos. O tamanho da população considerada foi de 100, 200, ou 400 indivíduos. Apenas um cromossomo com um único segmento de QTL segregante foi simulado com efeito aditivo, mas sem efeito de dominância. Os fatores aleatórios não genéticos seguiram distribuição normal com média zero e variância apropriada σ_e^2 . A variância aleatória não genética foi escolhida de modo que o QTL explicasse 1, 5, ou 10% da variância total. Para uma população de tamanho 100 e para uma variância com explicação de 1%, a simulação consistiu de 1.000 repetições, além de 500 repetições para os outros casos. O tamanho do cromossomo considerado foi de 120 cM, porém o tamanho médio de cromossomo utilizado em vários estudos em plantas é, geralmente, de 100 cM, considerando-se a ausência de interferência, com um marcador segregante a cada 5 cM, começando da posição 0, e as posições no mapa são assumidas como sendo precisamente conhecidas. Na prática, não são determinados todos os locos marcadores de uma só vez. Inicialmente, utilizam-se marcadores distanciados de 20 cM entre si, e, quando o LOD score tende a ser significativo, novos marcadores são adicionados. Entretanto, na simulação computacional feita neste estudo, todos os marcadores tiveram suas posições determinadas. A posição do QTL estava entre dois marcadores a 62,5 cM. Segundo o Lander e Botstein (1989), essa seria, presumidamente, a pior posição possível, uma vez que quanto mais próximo um QTL estiver de um dos marcadores, maior será o poder do mapeamento.

Um modelo para um único QTL foi usado. O LOD score foi calculado em posições a cada um cM. Quando o LOD score máximo para um cromossomo excede significativamente a entrada, intervalos-suporte são construídos com 0,5, 1, 2 e 3 LOD. A significância inicial é determinada previamente.

Ooijen (1992) comentou que o procedimento de mapeamento de QTL de Lander e botstein (1989) permite determinar a posição de QTLs em um mapa,

com limitações. QTLs com pequeno efeito aditivo ($\sigma_{\text{exp}}^2 = 1\%$) são muito difíceis de serem detectadas. O tamanho de uma população de pelo menos 200 indivíduos é necessário, a menos que se esteja interessado apenas em genes de efeito muito grande ($\sigma_{\text{exp}}^2 > 10\%$). O tamanho de população de 400 indivíduos parece ser o maior número possível praticamente em relação a trabalhos com RFLP. Contudo, pode ser esperado que com esse procedimento de mapeamento de QTL se explique uma variância de pelo menos 5% de chance de este ser detectado.

Em experimentos mais atuais de mapeamento, o mapa de ligação é sempre estimado através dos mesmos indivíduos, para os quais o mapeamento de QTL já foi feito. Tal mapa pode ser menos preciso, mas irá relacionar melhor eventos de recombinação nos cruzamentos.

Na prática, sempre haverá dados perdidos nos genótipos marcadores. Desde que esse fato leve a um menor LOD score e uma vez que a quantidade de dados perdidos vai variar dependendo do marcador, isso tem efeito na comparabilidade dos LOD scores de diferentes partes do cromossomo, a não ser que a quantidade não seja excessiva. De certa forma, isso também se aplica a marcadores dominantes. Estes levarão a um menor LOD score, e, conseqüentemente, será difícil a comparação de LOD scores baseados em dominantes com LOD scores baseados em marcadores co-dominantes.

Existem algumas etapas a serem seguidas para a construção de um mapa de ligação. Em geral, o primeiro passo na construção de um mapa de ligação está relacionado com a escolha dos genitores a serem cruzados, de forma que maximize o polimorfismo genético. Uma vez selecionados os genitores, é necessário o desenvolvimento de uma população segregante, composta de pouco mais de uma centena de indivíduos. É importante ressaltar que o número de marcadores polimórficos depende do polimorfismo genético entre os genitores e a precisão das estimativas de recombinação e cuidados em campo, entre outros, dependem da escolha da população (GRATAPAGLIA e SEDEROFF, 1994).

Após a escolha dos genitores e o desenvolvimento da população segregante, a etapa seguinte envolve a obtenção de marcas contrastantes entre os genitores que apresentam segregação mendeliana na população de mapeamento.

A estratégia de busca pelas marcas polimórficas depende, principalmente, do tipo de marcadores utilizados e da diversidade genética da espécie estudada (ROCHA, 2003).

Após a escolha dos marcadores polimórficos, é necessário analisar o padrão de amplificação dos indivíduos do restante de mapeamento. Na obtenção das estimativas de recombinação para construir um mapa genético, todos os marcadores devem ser analisados dois a dois, verificando-se a independência ou a existência de ligação entre eles (LIU, 1998). Com base no princípio de que os genótipos recombinantes alterados por permuta simples ou por permuta dupla são gerados em frequências diferentes, utiliza-se o teste de aderência χ^2 (qui-quadrado) para confirmar a ligação entre os marcadores. O teste de χ^2 é qualitativo e apenas comprova a existência ou não de ligação gênica. Permite calcular o desvio entre os resultados esperados, sem a ocorrência de permutas com os resultados observados, sendo sensível à magnitude do desvio e ao número de genótipos amostrados (FALCONER, 1987 *apud* ROCHA, 2003).

Confirmada a existência de ligação entre duas marcas, é indispensável adotar métodos quantitativos para estudar o grau de associação entre essas marcas. A metodologia de Máxima Verossimilhança é utilizada no mapeamento genético para a obtenção de várias estimativas, inclusive as da frequência de recombinação (LIU, 1998). Esse método permite a obtenção de estimadores consistentes, de distribuição normal, eficiência assintótica e variância mínima. A confiabilidade do posicionamento das marcas ao longo do grupo de ligação pode ser avaliada, considerando-se as variâncias associadas às estimativas de recombinação (LIU, 1998).

Ainda hoje, tanto o tamanho de população quanto o número de marcas para representação de cromossomos em grupos de ligação não são bem definidos, existindo falta de padrão para a análise de dados de trabalhos de mapeamento. Com as informações do Quadro 1, pode-se ter idéia da falta de padrão em relação ao número de indivíduos e de marcas a serem usadas em trabalhos com mapeamento genético.

Quadro 1 – Exemplos de trabalhos de diversos autores com espécies, tipo de população trabalhada, tamanho de população e número de marcadores diferentes

Espécie	Autor	Tipo de População	Tamanho	Nº de Marcadores
Bovinos	Telles (2001)	3 Rebanhos	66	133
Erva-mate	Vidor (2002)	Acessos	2.000	168
Eucalipto	Falcão (2004)	Parentais	8	1.261
Feijão	Corrêa (2000)	F ₂	302	440
Feijão	Faleiro (2003)	RILs	154	70
Maracujá	Ganga (2004)	Acessos	36	123
Milho	Brunelli (2002)	Linhagens, híbrido e F ₂ deles	165	142
Milho	Silva (2002)	F ₂	250	140
Morango	Conti (2002)	Cultivares	26	63
Simulado	Carbonell (1993)	Retrocruzamento e duplo haplóides	25.000	48
Simulado	Knott (1992)	F ₂	1.000	22
Soja	Corrêa (1999)	F ₁	50 e 120	238

Tamanho = número total de indivíduos analisados e N de marcadores = número de marcadores utilizados no genoma de cada indivíduo.

2.5. Simulação

A simulação consiste em construir um sistema que imite o funcionamento de uma realidade, com a finalidade de averiguar o que aconteceria no sistema real se alterações de interesse fossem efetuadas em seu funcionamento (DACHS, 1988). Informações valiosas podem ser extraídas desse sistema simulado, com menor custo e maior rapidez. No sistema real, a análise de muitas das opções de alteração é inviável, seja pelos custos que podem ser elevados, pelos longos períodos da resposta ou pela incerteza da direção e sentido das respostas, fatores que podem conduzir a um dano irreparável nesse sistema.

No melhoramento, o tempo se torna fator limitante. Além do tempo necessário, a necessidade de laboratório aumenta os custos, dificultando, muitas vezes, a realização de determinados trabalhos de pesquisa. Para contornar esse problema, pesquisadores utilizam técnicas de simulação que permitem a obtenção de grande volume de dados em curto período de tempo, sem os custos de implantação e condução de experimentos com animais ou plantas e com laboratórios (CORRÊA, 2001).

Ao fazer estudos baseados em simulação, deve-se considerar que nada é tão simples que possa ser compreendido e controlado sem abstração. Essa abstração consiste em substituir o objeto de interesse por um modelo semelhante, porém com estrutura mais simples. Outro aspecto importante na simulação é a modelagem. O modelo deve ser suficientemente simples para ser operacionalizado e interpretado adequadamente, mas seu desempenho deve ser comparável com o modelo real e, se a defasagem for grande, ele deve ser eliminado ou refinado (CRUZ *apud* NASS, 2001). Ao utilizar uma técnica de simulação, o pesquisador deve se precaver contra erros, sejam estes devidos a problemas com levantamentos amostrais, escolha inadequada das distribuições de probabilidades nos eventos de natureza aleatória, simplificação inadequada da realidade e erros de implantação do sistema simulado. Para a garantia de sua eficiência, pode-se lançar mão de processos de validação. Essa validação consiste em fazer o sistema simulado operar nas condições do sistema real e verificar, através de testes de hipóteses e outras análises estatísticas ou através de comparação com situações reais já analisadas, se os resultados observados na simulação condizem com os constatados no sistema real (CRUZ *apud* NASS, 2001).

Dempster et al. (1977) apresentaram uma abordagem de um método computacional iterativo para estimções de máxima verossimilhança, quando as observações são classificadas como dados incompletos. A partir do momento em que cada iteração do algoritmo consiste em uma etapa de estimativa e outra de maximização, esse algoritmo é chamado de EM. O algoritmo tem ampla aplicabilidade, incluindo dados perdidos, dados truncados, modelos de “misturas finitas” e estimação de componentes de variância, análise de fatores e outros, além de ser relativamente simples.

No trabalho de Frish (1999) foi usado o software PLABSIM, que é um programa computacional desenvolvido em C++, para simular o processo de recombinação, durante o processo de meiose. Os eventos de permutação foram gerados. As frequências de recombinação necessárias para o processo de simulação de permuta foram calculadas pelas distâncias de mapa; pela função de

mapeamento de Haldane (1919), isso foi feito assumindo-se que não existe interferência de quiasmas.

Darvasi et al. (1993) realizaram estudo de simulação, usando populações de retrocruzamento, para determinar o efeito do espaçamento de marcadores, o efeito do gene e o tamanho da população no poder da ligação entre o marcador e o QTL, no erro-padrão dos estimadores de máxima verossimilhança e do efeito do QTL sobre a localização no mapa de ligação. Esse desenho experimental foi escolhido por sua simplicidade analítica e ampla utilização na prática. Neste estudo, constatou-se que o poder de detecção de QTL foi virtualmente o mesmo, usando marcadores espaçados de 10 ou um número infinito de marcadores, com a ocorrência de pequeno decréscimo quando os marcadores foram espaçados de 20 ou até mesmo de 50 cM. A vantagem de utilizar intervalo de mapeamento em vez de análise com apenas um marcador foi sem importância. O poder de resolução de uma ligação marcador-QTL foi definido como sendo de 95% do intervalo de confiança, para um mapa de localização de QTL que seria obtido através da utilização de infinitos marcadores. Descobriu-se que, reduzindo o espaçamento entre marcadores para menos do que o poder de resolução, não houve ganho apreciável estreitando o intervalo de confiança. É interessante observar que, nesse trabalho de simulação, foram geradas 1.000 repetições para cada parâmetro de combinação.

O método de máxima verossimilhança para o mapeamento de QTL também tem sido estudado em populações F_2 usando-se dados simulados. Métodos de máxima verossimilhança podem ser computacionalmente exigentes, porém apresentam vantagens por possibilitar testar hipóteses e estimação de parâmetros simultaneamente, além de permitir ajustar o modelo de modo que possa levar em consideração e estimar a fração de recombinação. Os métodos de máxima verossimilhança são muito versáteis, permitindo que se regule o modelo para que diferentes valores de variâncias residuais se ajustem a diferentes genótipos de QTL. É visto que o uso de pares de marcadores flanqueadores aumenta o poder de detecção de QTL em relação ao uso de apenas um marcador, quando os marcadores estão muito espaçados e o efeito do QTL é grande. O uso

de marcadores flanqueadores também proporciona estimativas mais acuradas e com menor viés para a localização e efeito do QTL. Além disso, o método se torna menos sensível a violações ou pressuposições, por exemplo uma distribuição não normal. Ao serem testadas hipóteses de um QTL ligado com a de não existência de QTL, não houve viés devido à presença de QTL não ligado. Ao ser fixada a fração de recombinação entre os marcadores a um valor incorreto, uma análise com marcadores flanqueadores geralmente não gera um viés no teste para estimação do QTL ou do seu efeito. Porém, a existência de muitos QTLs ligados provoca viés em testes e nas estimativas do efeito do QTL com mapeamento por intervalo, levando a valores maiores, ou inflados, quando QTLs estão associados às linhas cruzadas e a valores menores quando estão em dispersão (KNOTT e HALEY, 1992).

Carbonell et al. (1993), em estudos visando à estimação de parâmetros genéticos e à localização de QTLs para populações de retrocruzamento e duplo haplóides, utilizaram 100 conjuntos de 250 indivíduos simulados, tendo oito grupos de ligação com seis marcadores cada, distanciados de 20 cM. Foram usados seis QTLs envolvidos na expressão de uma característica, sendo que alguns genes expressavam efeitos de dominância e outros, aditividade completa. Os estudos foram feitos considerando-se sistema de cruzamentos para populações de retrocruzamento e duplo haplóides. Os parâmetros da simulação foram estabelecidos de modo a refletir diferentes condições experimentais, permitindo, assim, a realização de comparações de interesse. É importante ressaltar que o tamanho das amostras foi mantido pequeno para permitir que aparecessem diferenças nas comparações realizadas. Os autores concluíram que populações de duplo haplóides permitem que experimentos possam ter menor tamanho de suas amostras por possuírem maior poder que populações de retrocruzamento. Além disso, populações de duplo haplóides possibilitam estimações mais acuradas da localização do QTL, com menor variância.

Ainda segundo esses autores, quando vários QTLs estão localizados no mesmo grupo de ligação, os resultados de um único experimento podem ser enganosos. Pode acontecer de o pesquisador inferir sobre a existência de um

único QTL de grande efeito, que na verdade está localizado entre outros dois QTLs. Isso tem maior probabilidade de acontecer quando se testa uma amostra pequena e a herdabilidade é baixa. Na maioria das vezes, o pesquisador pode encontrar um platô ou uma distribuição multimodal, sendo essas distribuições o indicativo de que vários QTLs podem estar envolvidos, por isso um estudo mais detalhado deve ser feito.

Martinez et al. (1999) desenvolveram procedimento de mapeamento baseado em modelo randômico utilizado para investigar sua robustez e adequação para mapeamento de QTL em populações em que prevalecem estruturas de famílias de meios-irmãos. Sob o modelo randômico, a localização do QTL e componentes de variância foram estimados usando-se técnicas de máxima verossimilhança. A estimação de parâmetros é feita com base na abordagem do modelo de pares de irmãos. A proporção de genes idênticos por descendência (IBD) no QTL foi estimada através de dois marcadores flanqueadores. Já as estimativas dos parâmetros e do poder de QTL foram obtidos usando-se dados simulados, variando o número de famílias analisadas, a herdabilidade da característica, a variância, o número de marcadores e o número de alelos no QTL. Os fatores mais importantes que influenciam o poder e os parâmetros do QTL foram a herdabilidade e a variância genética da característica. Segundo os referidos autores, o número alelos do QTL não influenciou as estimativas dos parâmetros avaliados nem o poder de detecção do QTL. Com uma herdabilidade mais alta, foi observado um confundimento entre QTL e componentes poligênicos.

Em relação ao trabalho de Martinez et al. (1999), a técnica de simulação Monte Carlo foi utilizada para gerar dados genotípicos e fenotípicos. O mapeamento de QTL foi considerado para um segmento de cromossomo de 100 cM de comprimento, coberto por seis marcadores distribuídos igualmente ao longo do cromossomo, a uma distância de 20 cM entre eles. Todos os marcadores tinham igual número de alelos e de mesma frequência. Um único QTL com vários alelos co-dominantes de mesma frequência e de efeito aditivo foi simulado no meio do segmento cromossômico, 50 cM, e os pais foram gerados através da

alocação aleatória dos genótipos em cada loco, assumindo-se equilíbrio de Hardy-Weinberg. A fase de ligação parental foi assumida como sendo desconhecida. Descendentes foram gerados assumindo-se não haver interferência. Desse modo, o evento de recombinação em um intervalo não irá interferir num evento de recombinação em um intervalo adjacente. As frações de recombinação, em cada loco, foram calculadas usando-se a função de mapeamento de Haldane. É importante mencionar que, em cada simulação, dois tamanhos diferentes de amostras foram considerados (50 e 100 famílias) de pais com 25 descendentes cada.

Para testar o comportamento do modelo randômico aproximado sob diferentes valores de herdabilidade das características e diferentes proporções da variância explicada pelo QTL, sete diferentes valores de herdabilidade foram assumidos: a herdabilidade da característica variou de 0,10 a 0,70, em passos de 0,10. A variância genética total consistiu de um componente de QTL e de um componente poligênico não ligado. O efeito aditivo do alelo do QTL foi dado de forma que a variância do QTL fosse de 10, 50 e 100% do total da variância genética. Todos os seis marcadores tinham seis alelos, todos com a mesma frequência. Para testar a influência dos polimorfismos dos marcadores no desempenho do modelo randômico aproximado, assumiu-se que cada um dos seis marcadores teria dois, quatro, seis ou 10 alelos com frequências iguais. Foram considerados dois valores diferentes de herdabilidade, 0,10 e 0,50. O número de alelos no QTL era 5. A variância genética total foi atribuída ao QTL, não sendo simulado nenhum componente poligênico.

2.6. Contribuição dos estudos de simulação na análise genômica

Direta ou indiretamente, a simulação tem contribuído grandemente para o avanço tanto na Genômica quanto nas demais áreas do melhoramento.

Uma dessas contribuições é a possibilidade de serem simulados diferentes tipos de populações, permitindo menor gasto em material e mão-de-obra, além do ganho de tempo. As simulações podem também ser feitas com base em dados

reais. Frish, et al. (1999), para verificar a confiabilidade do software utilizado, compararam o mapa de ligação original utilizado em seu trabalho, no qual foi baseado em dados experimentais de F_2 , com um mapa de ligação construído a partir de dados simulados de indivíduos F_2 , pelo programa MAPMAKER (LANDER et al., 1987). Pela comparação, foi visto que os mapas estavam em excelente concordância, confirmando que os modelos usados nos dois programas eram similares.

Mais uma vantagem do uso da simulação pode ser encontrada no trabalho de Visscher (1996), em que a simulação foi utilizada para avaliar eficiência do uso da seleção assistida na introgressão de genes em programas de retrocruzamento. No trabalho de Martínez (1992), também foram feitas simulações correspondentes a um retrocruzamento, para verificar, dentre outros, o efeito de QTL.

Outra vantagem da simulação é que ela pode ser empregada em qualquer etapa de um programa de melhoramento, seja este clássico ou não, para comprovar ou refutar novos procedimentos e técnicas. Uma das formas de simulação vista ao longo dos textos anteriores foi a utilização da simulação para obtenção de dados sobre determinadas populações. Desse modo, após chegar a um modelo para detecção de QTL, por exemplo, pode-se testar sua eficiência em populações de retrocruzamento, F_1 , F_2 , duplo-haplóides, linhas endogâmicas e outras, sem o custo operacional e laboratorial que poderia ser gerado.

Para serem simuladas populações de diversos tipos de famílias, podem ser usadas características obtidas de dados reais, como tipo de distribuição, média, variância e demais características que se mostrarem necessárias.

Uma clara vantagem do uso da simulação é o número de amostras que podem ser geradas. Como exemplo, pode-se citar o trabalho de Martínez (1992), em que, para ilustrar como aparece um QTL fantasma quando é usado o mapeamento por intervalo de Lander e Botstein (1989) e o mapeamento por regressão usando apenas marcadores flanqueadores, foi selecionada uma amostra com 2.000 observações, sendo esse número de observações escolhido por se querer observar melhor o viés sistemático, em vez da variação de amostra.

Em seu trabalho sobre técnicas de máxima verossimilhança para mapeamento e análise de QTL com auxílio de marcadores genéticos, Weller (1986), citado por Carbonell (1993), apresentou dados simulados que indicaram boa concordância entre os efeitos preditos e os obtidos em 10 repetições de 2.000 indivíduos F_2 . Pelo comentado, a técnica se mostrou mais eficiente para genes co-dominantes que para dominantes.

No trabalho de Knott e Haley (1992) para investigar as propriedades dos métodos analíticos, empregaram-se dados simulados. Foram gerados descendentes de parentais completamente heterozigotos, assumindo-se que não existam interferência e aditividade em Morgan das distâncias de mapa. Cada conjunto de dados continha 1.000 indivíduos F_2 , e o genótipo de cada indivíduo foi composto por um par de cromossomos de 100 cM de comprimento. Simularam-se 11 locos marcadores, em um intervalo de 10 cM entre eles. Também foram simuladas diferentes situações de ligação de QTLs.

Uma das dificuldades atualmente enfrentadas pelos pesquisadores é a identificação da posição de QTLs e do tamanho do seu efeito. Através da simulação, pode-se inferir e testar, sob determinadas pressuposições, o que aconteceria sob certas situações com um maior distanciamento entre marcas moleculares, qual seria o efeito da presença de um ou mais QTLs próximos a um QTL flanqueado por marcadores, o que aconteceria caso existissem QTLs externamente em um intervalo limitado por marcadores, porém sem a existência de QTL entre eles, e outras situações que se mostrassem de interesse para análise. Martínez (1992), procurando estimar a localização e o tamanho dos efeitos de QTLs usando marcadores flanqueadores, gerou um conjunto de dados simulados que exemplificaram o problema a ser pesquisado, sendo os dados analisados pelo método de mapeamento, por intervalo e pelo modelo de regressão.

2.7. Programas GENES e GQMOL

O programa GENES, amplamente utilizado em análises de modelos aplicados ao melhoramento de plantas e animais, é um software destinado à

análise e processamento de dados por meio de diferentes modelos biométricos, contando com procedimentos uni e multivariados e enfatizando a estimação de parâmetros genéticos. Também estão disponíveis procedimentos para análise de dados binários, geralmente obtidos de estudos moleculares, permitindo a análise e interpretação de fenômenos particulares dessa área (CRUZ, 2004).

O programa GENES está disponível para download, gratuitamente, sem nenhuma restrição de uso ou de divulgação, no endereço <<http://www.ufv.br/dbg/genes/genes.htm>>, e conta com manual (livro) comercializado pela Editora UFV: editora@ufv.br ou e <<http://www.livraria.ufv.br/>>.

O programa GQMOL foi desenvolvido com o propósito de analisar dados obtidos de estudos moleculares. Pode ser usado em análise de segregação de locos individuais, estimação de porcentagem de recombinação, agrupamento de marcas moleculares e mapeamento, incluindo estudos de QTL com populações controladas, populações exogâmicas, simulação e análise de imagem. Além disso, conta com um módulo de ensino, apresentando vários procedimentos para entendimento de princípios estatísticos e genéticos envolvidos na análise genômica.

O programa GQMOL está disponível para download, gratuitamente, sem nenhuma restrição de uso ou de divulgação, no endereço <<http://www.ufv.br/dbg/gqmol/gqmol.htm>>.

3. MATERIAL E MÉTODOS

3.1. Simulação de dados

Para gerar os dados foi utilizado o módulo de simulação do aplicativo computacional GQMOL (CRUZ, 2005), o qual permite gerar informações sobre genomas, genótipos, genitores, indivíduos de diferentes tipos de populações e dados de características quantitativas. Foram simulados genomas parentais e amostras de populações F_2 e de retrocruzamento. As amostras geradas foram de 50, 100, 150, 200, 400, 600 e 800 indivíduos, com 10 grupos de ligação cada. Também foram geradas matrizes de distâncias entre pares de locos. Após a simulação e geração da matriz de distância, foi avaliada a eficácia de reconstituição do número de grupos de ligação a partir dos diferentes tamanhos e tipos de amostras.

3.1.1. Simulação do genoma

Para este estudo foi tomada como referência uma espécie diplóide fictícia com $2n = 2x = 20$ cromossomos, cujo comprimento total do genoma foi estipulado em 1.000 cM. Foram gerados genomas com níveis de saturação de 5, 10, 15 e 20 marcas moleculares (ou 20, 10, 6,66 e 5 cM, respectivamente) por

grupo de ligação. Cada genoma foi composto por 10 grupos de ligação, 100 cM em cada grupo, com um comprimento total de 1.000 cM.

Neste estudo foram usadas nas simulações marcas co-dominantes para retrocruzamento e populações F_2 em saturações com 5, 10, 15 e 20 marcas por grupo de ligação, como pode ser observado na Figura 1.

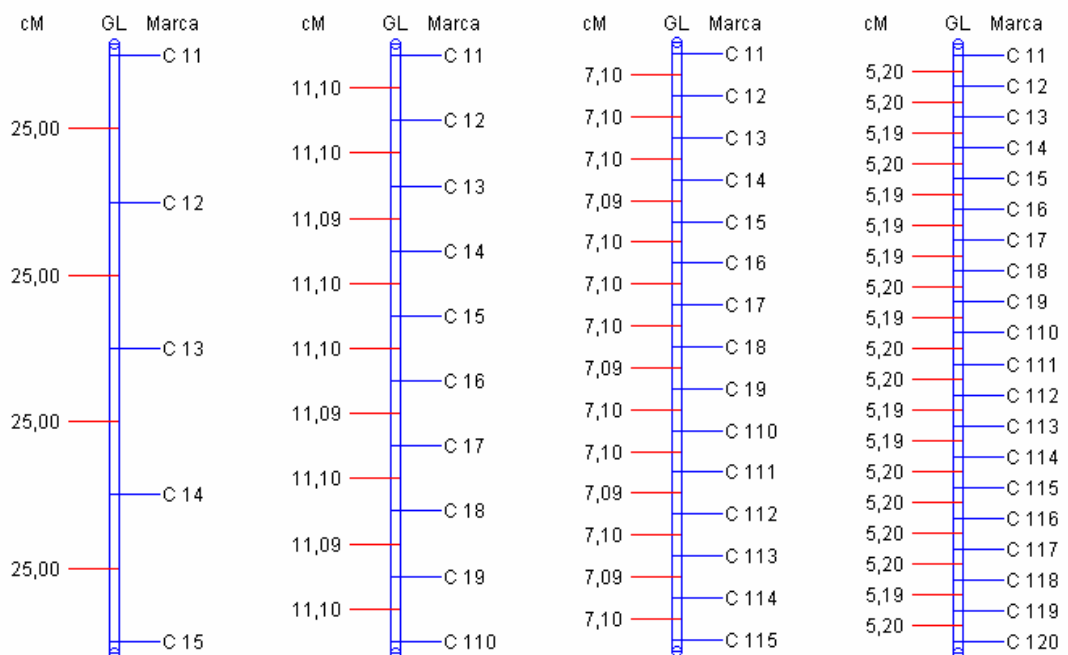


Figura 1 – Grupo de ligação 1 com 5, 10, 15 e 20 marcadores, evidenciando-se os diferentes graus de saturação.

3.1.2 Simulação dos genitores

Na análise de ligação gênica com dados de populações F_2 e de retrocruzamento, para cada grau de saturação do genoma estudado foi simulada a situação em que os pais são homozigotos, sendo um dominante e outro recessivo. Também foi concebido que todos os locos estivessem em aproximação. A seguir é exemplificado o genótipo dos pais P_1 e P_2 em diferentes graus de saturação (Figura 2), referentes ao primeiro grupo de ligação (GL1).

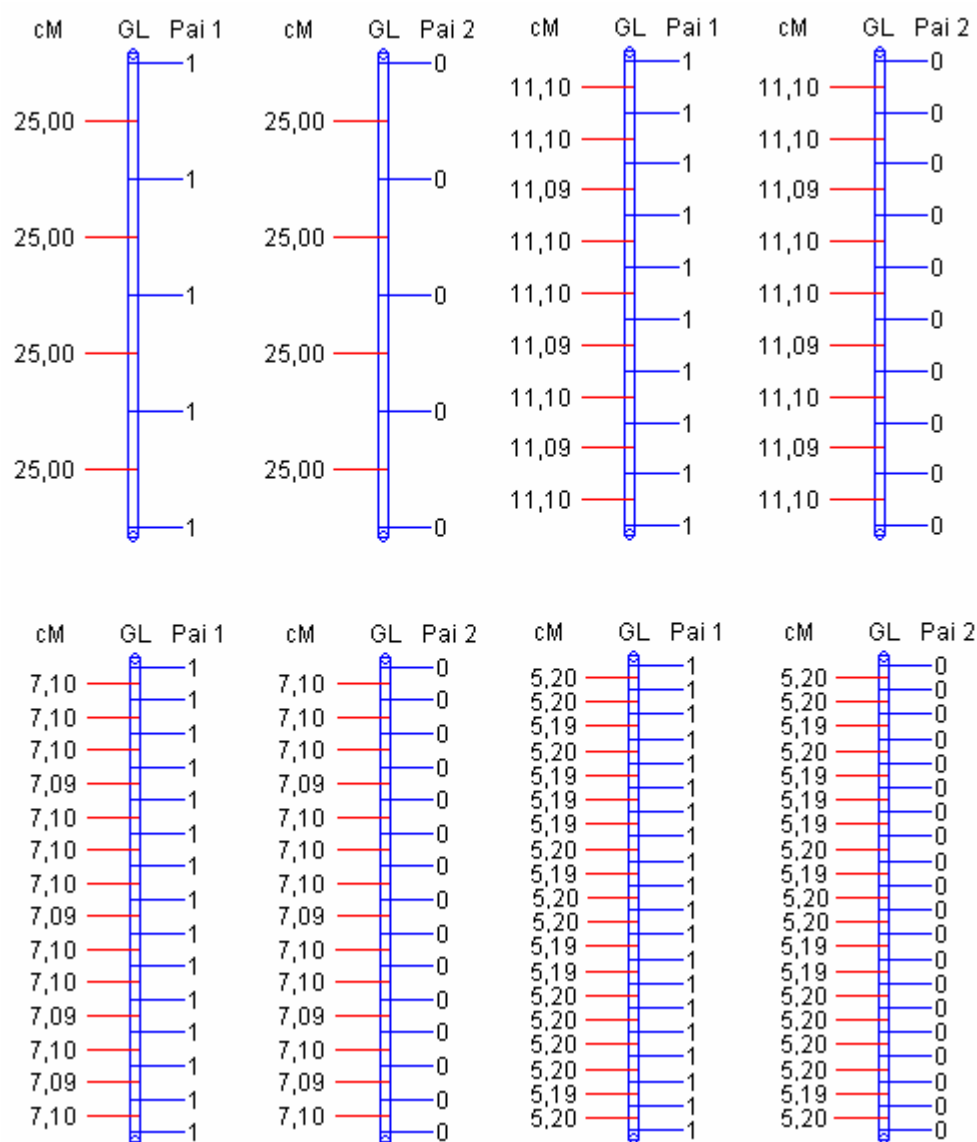


Figura 2 – Genoma dos pais 1 e 2, com quatro níveis de saturação. O código 1 representa a presença de marca e o código 0, a ausência em determinado loco.

3.1.3. Tamanho da população

Neste trabalho, tanto para a população de retrocruzamento quanto para as populações F_2 foram geradas amostras com 50, 100, 150, 200, 400, 600 e 800 indivíduos, com saturações de 5, 10, 15 e 20 marcas por grupo de ligação e 100 repetições para cada nível de saturação, chegando a um total de 5.600 simulações, sendo 2.800 para retrocruzamento e 2.800 para população F_2 .

3.1.4. Procedimento de simulação dos indivíduos das populações F_2 e de retrocruzamento

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 e de retrocruzamento seguiu os seguintes passos:

- i) A partir do genoma simulado, foram construídos os genótipos parentais homozigotos e contrastantes para os marcadores, de tal forma que a geração F_1 estivesse em aproximação para todos os marcadores.
- ii) A partir do genótipo da geração F_1 , foram gerados os gametas para a formação dos indivíduos das populações F_2 e de retrocruzamento. A produção de gametas foi feita simulando-se o pareamento dos homólogos e realizando-se permutas ao longo dos cromossomos e considerando a não-existência de interferência nas regiões delimitadas por dois marcadores adjacentes. A probabilidade de ocorrência de recombinação numa região entre marcadores adjacentes foi dada de acordo com a distância desses marcadores no genoma simulado. Ressalta-se que maior distância implica maior possibilidade de ocorrência de recombinação.

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

3.2. Análise genômica – Mapeamento

Após a geração dos dados, seguiram-se as etapas do processo de mapeamento, como descrito nos tópicos subseqüentes.

3.2.1. Análise de segregação de locos individuais

Foram aplicados testes de qui-quadrado (χ^2) para verificação da razão de segregação em cada marca de todas as populações geradas. No processo de mapeamento foram utilizadas todas as marcas, mesmo as que não segregaram de acordo com a proporção esperada de 1:1 (Aa:aa) para retrocruzamento e 1:2:1 (aa:Aa:AA) para marcas co-dominantes.

A estatística de qui-quadrado é dada por:

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

em que:

χ^2 = valor de qui-quadrado calculado; e

Obs_i e Esp_i = valores observados e esperados, na i-ésima classe fenotípica (i = 1, 2, ..., n), respectivamente.

A hipótese (H₀) de segregação dos locos foi testada a 5% de probabilidade. Nas situações em que o valor da probabilidade calculado foi inferior ao preestabelecido, a hipótese H₀ foi rejeitada, significando que a segregação não ocorreu de acordo com o esperado.

3.2.2. Estimação da porcentagem de recombinação

Após a aplicação dos testes de segregação, seguiu-se a etapa da estimativa da porcentagem de recombinação entre pares de marcas, utilizando o método da máxima verossimilhança.

3.2.2.1. População de retrocruzamento

A descendência do retrocruzamento foi estabelecida a partir do cruzamento do F_1 com o genitor homocigoto recessivo, de forma que a frequência genotípica esperada para a progênie desse retrocruzamento fosse igual à frequência gamética do F_1 , pois o genitor homocigoto recessivo produzirá gametas de um único tipo (Quadro 2).

Quadro 2 – Frequências genotípicas esperadas para a progênie do retrocruzamento entre o F_1 e o genitor aabb

Classes	Ocorrência	Frequência	Tipo
AaBb	N_1	$1/2(1-r)$	Paternal
aabb	N_2	$1/2(1-r)$	Paternal
Aabb	N_3	$r/2$	Recombinante
aaBb	N_4	$r/2$	Recombinante

A estimação da porcentagem de recombinação pode ser feita pelo método da máxima verossimilhança, levando-se em consideração a função de verossimilhança:

$$L(r; n_i) = \lambda \left[\frac{1}{2}(1-r) \right]^{n_1} \left[\frac{1}{2}(1-r) \right]^{n_2} \left(\frac{1}{2}r \right)^{n_3} \left(\frac{1}{2}r \right)^{n_4}$$

em que $\lambda = \frac{N!}{n_1!n_2!n_3!n_4!}$, sendo $N = n_1+n_2+n_3+n_4$.

A função suporte é dada por:

$$\lambda(r; n_i) = \ln(\lambda) + N \ln\left(\frac{1}{2}\right) + (n_1 + n_2) \ln(1-r) + (n_3 + n_4) \ln(r)$$

Assim, por meio de função escore, obtida pela primeira derivada da função suporte, é possível chegar ao estimador da frequência de recombinação:

$$\frac{\partial \lambda(r; n_i)}{\partial r} = \frac{(n_1 + n_2)}{1-r}(-1) + \frac{(n_3 + n_4)}{r} = \frac{(n_3 + n_4) - Nr}{r(1-r)} = f'(r)$$

igualando a 0, obtém-se o estimador:

$$r = \frac{n_3 + n_4}{N}$$

Esse estimador também expressa o conceito de distância entre locos gênicos dado pela frequência de gametas recombinantes.

Para que seja obtida a variância da estimativa da porcentagem de recombinação, considera-se a função $f'(r)$, de modo que $H(r) = r(1-r)$ e $G(r) = (n_3 + n_4) - Nr$. Com isso, tem-se que $G'(r) = -N$, de forma que o índice de informatividade será:

$$I(r) = -f''(r) = \frac{N}{r(1-r)}$$

O conteúdo médio de informação, uma característica da população, é expresso, para retrocruzamento, por meio de:

$$c(r) = \frac{1}{r(1-r)}$$

3.2.2.2. População F₂ com marcadores co-dominantes

A geração F₂ é obtida pela autofecundação dos indivíduos F₁. Em se tratando de um par de marcadores, as seguintes combinações são possíveis: apenas marcadores co-dominantes, apenas marcadores dominantes e marcadores dominantes e co-dominantes simultaneamente. Em qualquer uma dessas situações deve ser considerada a frequência esperada dos genótipos na geração F₂ (Quadro 3).

Quadro 3 – Frequências genômicas esperadas na população F_2 , utilizando-se marcadores co-dominantes, derivados do cruzamento entre indivíduos AABB x aabb

Genótipo	Observação	Frequência
AABB	n_1	$p_1 = 0,25(1-r)^2$
AABb	n_2	$p_2 = 0,50r(1-r)$
AAbb	n_3	$p_3 = 0,25r^2$
AaBB	n_4	$p_4 = 0,50r(1-r)$
AaBb	n_5	$p_5 = 0,50(1-r)^2 + 0,50r^2$
Aabb	n_6	$p_6 = 0,50r(1-r)$
aaBB	n_7	$p_7 = 0,25r^2$
aaBb	n_8	$p_8 = 0,50r(1-r)$
aabb	n_9	$p_9 = 0,25(1-r)^2$

$P_i(R|G)$ é a probabilidade de que um gameta seja recombinante, dado o genótipo dos dois locos.

Neste trabalho foram consideradas apenas as situações que envolviam marcadores co-dominantes, considerando-se que:

$$n_a = n_1 + n_9$$

$$n_b = n_2 + n_4 + n_6 + n_8$$

$$n_c = n_3 + n_7$$

$$n_d = n_5$$

$$N = n_a + n_b + n_c + n_d$$

Para estimar as frequências de recombinação, a função de máxima verossimilhança é dada por:

$$L(r; n_i) = \lambda p_1^{n_1} p_2^{n_2} \dots p_9^{n_9}$$

sendo:

$$\lambda = \frac{N!}{n_1! n_2! \dots n_9!}$$

ou

$$L(r; n_i) = \lambda \left[\frac{1}{4}(1-r)^2 \right]^{n_a} \left[\frac{1}{2}r(1-r) \right]^{n_b} \left[\frac{1}{4}r^2 \right]^{n_c} \left[\frac{1}{2}(1-r)^2 + \frac{1}{2}r^2 \right]^{n_d} =$$

$$L(r; n_i) = \lambda \alpha_1^{n_a} \lambda \alpha_2^{n_b} \lambda \alpha_3^{n_c} \lambda \alpha_4^{n_d}$$

Tem-se, portanto, a função suporte:

$$\lambda(r; n_i) = \ln(\lambda) + n_a \ln(\alpha_1) + n_b \ln(\alpha_2) + n_c \ln(\alpha_3) + n_d \ln(\alpha_4)$$

Por derivação da função suporte em relação à r , tem-se a função escore dada por:

$$\frac{\partial \lambda(r; n_i)}{\partial r} = \frac{n_a}{(1-r)^2} 2(1-r)(-1) + \frac{n_b}{r(1-r)} (1-2r) + \frac{n_c}{r^2} 2r + \frac{n_d}{(1-r)^2} [2(1-r)(-1) + 2r]$$

de forma que:

$$\frac{\partial \lambda(r; n_i)}{\partial r} = \frac{(n_b + 2n_c) - 2r(n_a + n_b + n_c)}{r(1-r)} + \frac{2(2r-1)n_d}{(1-r)^2 + r^2}$$

igualando a 0 a função escore, tem-se:

$$(n_b + 2n_c) - 2(n_a + 2n_b + 3n_c + n_d)r + 2(2n_a + 3n_b + 4n_c + 3n_d)r^2 - 4Nr^3 = 0$$

Assim, a solução de máxima verossimilhança é obtida por uma das raízes da equação polinomial de terceiro grau.

Para obter a variância da estimativa, considera-se a função suporte:

$$f'(r) = \frac{(n_b + 2n_c) - 2(n_a + 2n_b + 3n_c + n_d)r + 2(2n_a + 3n_b + 4n_c + 3n_d)r^2 - 4Nr^3}{r(1-r)(1-2r+2r^2)}$$

portanto, adotando as funções $G(r)$ e $H(r)$, obtém-se:

$$G(r) = (n_b + 2n_c) - 2(n_a + 2n_b + 3n_c + n_d)r + 2(2n_a + 3n_b + 4n_c + 3n_d)r^2 - 4Nr^3$$

e

$$H(r) = r(1-r)(1-2r+2r^2)$$

logo:

$$-G'(r) = 2(n_a + 2n_b + 3n_c + n_d) - 4(2n_a + 3n_b + 4n_c + 3n_d)r + 12Nr^2$$

sabe-se que:

$$E(n_a) = \left[\frac{1}{2}(1-r)^2 \right] N$$

$$E(n_b) = 2r(1-r)N$$

$$E(n_c) = \frac{1}{2}r^2N$$

e

$$E(n_d) = \left[\frac{1}{2}(1-2r+r^2) + \frac{1}{2}r^2 \right] N$$

então:

$$-G'(r) = 2N(1-3r+3r^2)$$

e, assim:

$$I(r) = \frac{2(1-3r+3r^2)N}{r(1-r)(1-2r+2r^2)}$$

O conteúdo médio de informação, para a população F_2 e marcadores co-dominantes, é expresso por meio de:

$$c(r) = \frac{2(1-3r+3r^2)}{r(1-r)(1-2r+2r^2)}$$

3.3. Comparação de genomas

O termo genoma simulado refere-se, genericamente, àqueles obtidos conforme descrito na seção 3.1.1 do item Material e Métodos. Já o termo genoma analisado se refere, genericamente, aos genomas construídos a partir das populações simuladas.

Foram comparados, nos genomas analisados, o número de grupos de ligação obtidos, o tamanho dos grupos de ligação, as distâncias médias entre dois marcadores adjacentes nos grupos de ligação, as variâncias das distâncias entre

marcas adjacentes nos grupos de ligação, o estresse e se houve ou não inversão da ordem dos marcadores, verificada pela correlação de Spearman. Todas essas comparações foram realizadas com o módulo “Comparação de genomas” do aplicativo computacional GQMOL (CRUZ, 2005).

Nas análises apresentadas foram utilizadas apenas as repetições em que houve recuperação de 10 grupos de ligação no mapeamento genético. Exceção feita somente nas amostras em que foram utilizadas todas as 100 repetições, independentemente do número de grupos de ligação formados.

3.3.1. Número de grupos de ligação e marcas por grupo

Em todos os genomas analisados, fez-se uma contagem do número de grupos de ligação e do número de marcas por grupo, obtidos do mapeamento das populações simuladas.

3.3.2. Tamanho do grupo de ligação

O tamanho do grupo de ligação foi obtido, somando-se as distâncias entre marcas adjacentes no grupo de ligação do genoma analisado, como segue:

$$L = \sum_{k=1}^{m-1} d_k$$

em que L é o tamanho do grupo de ligação e d_k , a distância entre marcas adjacentes m_k e m_{k+1} no grupo de ligação do genoma analisado ($k = 1, \dots, m-1$). Sendo que m é o número de marcadores no grupo de ligação do genoma analisado.

3.3.3. Média das distâncias entre dois marcadores adjacentes no grupo de ligação

É a razão do tamanho do grupo de ligação pelo número de intervalos entre marcas adjacentes no grupo de ligação, como segue:

$$\bar{d} = \frac{L}{I}$$

em que \bar{d} é a distância média de dois marcadores adjacentes no grupo de ligação do genoma analisado, L é o tamanho do grupo de ligação do genoma analisado e I é o número de intervalos entre marcas adjacentes, dado por $m - 1$, sendo m o número de marcas no grupo de ligação.

3.3.4. Variância das distâncias entre marcas adjacentes

É a razão do somatório do quadrado dos desvios entre as distâncias de marcas adjacentes e a distância média de dois marcadores adjacentes no grupo de ligação pelo número de intervalos (I) no grupo de ligação menos 1, como segue:

$$\hat{\sigma}^2 = \frac{\sum_{k=1}^{m-1} (d_k - \bar{d})^2}{I - 1}$$

em que $\hat{\sigma}^2$ é a variância das distâncias entre marcas adjacentes, d_k é a distância entre marcas adjacentes m_k e m_{k+1} no grupo de ligação do genoma analisado ($k=1, \dots, m-1$), \bar{d} é a média da distância de dois marcadores adjacentes no grupo de ligação do genoma analisado e I é o número de intervalos entre marcas adjacentes, dado por $m-1$, sendo m o número de marcadores no grupo de ligação.

3.3.5. Correlação de Spearman

A correlação de spearman, também conhecida como correlação de rank, é utilizada quando não é possível mensurar variações contínuas, como variáveis x e y nos n membros de uma população. Porém, é possível mensurar um x e um y , em forma de nota (rank), em que cada nota pode ser colocada em ordem nos n membros. Essa correlação expressa o grau de concordância nas notas das duas variáveis.

Adaptando a correlação de Spearman, dada por Clarke (1994), para análise de genomas, tem-se:

$$r_s = 1 - \frac{6 \sum_{k=1}^m \Delta_k^2}{m(m^2 - 1)}$$

em que r_s é o valor estimado da correlação de Spearman, para um grupo de ligação do genoma analisado ($-1 \leq r_s \leq 1$), Δ_k é a diferença da nota do marcador m_k ($k=1, \dots, m$) na posição k do grupo de ligação do genoma simulado e a posição do marcador m_k na posição k do grupo de ligação do genoma analisado, em que m é o número de marcas no grupo de ligação do genoma simulado e a nota do marcador m_k , tanto no grupo de ligação do genoma simulado quanto no grupo de ligação do analisado, é o valor do índice k do referido marcador. O valor da nota do marcador não é alterado quando sua posição em relação ao grupo de ligação do genoma simulado é alterada. Deve-se ressaltar que a nota se refere à posição que determinado marcador ocupa no grupo de ligação de referência e não ao valor ou importância do marcador. Por exemplo (Figura 3), se no grupo de ligação do genoma simulado a ordem dos marcadores for m_1 - m_2 - m_3 -...- m_9 , então a nota do marcador m_1 será 1, a do m_2 será 2, a do m_3 3.... E, se no grupo de ligação do genoma analisado a ordem dos respectivos marcadores for m_2 - m_1 - m_3 -...- m_9 , a nota do marcador m_2 continuará sendo 2, do m_1 será 1, do m_3 será 3 e até o último marcador.

Portanto, os valores de Δ_k serão:

$$\Delta_1 = (1-2) = -1, \Delta_2 = (2-1) = 1, \Delta_3 = (3-3) = 0 \dots \Delta_9 = (9-9) = 0$$

Com isso, o valor estimado da correlação de Spearman é:

$$r_s = 1 - \frac{6[(-1)^2 + (1)^2 + (0)^2 + (0)^2 + (0)^2 + (0)^2 + (0)^2 + (0)^2 + (0)^2]}{9(9^2 - 1)} = 0,9833$$

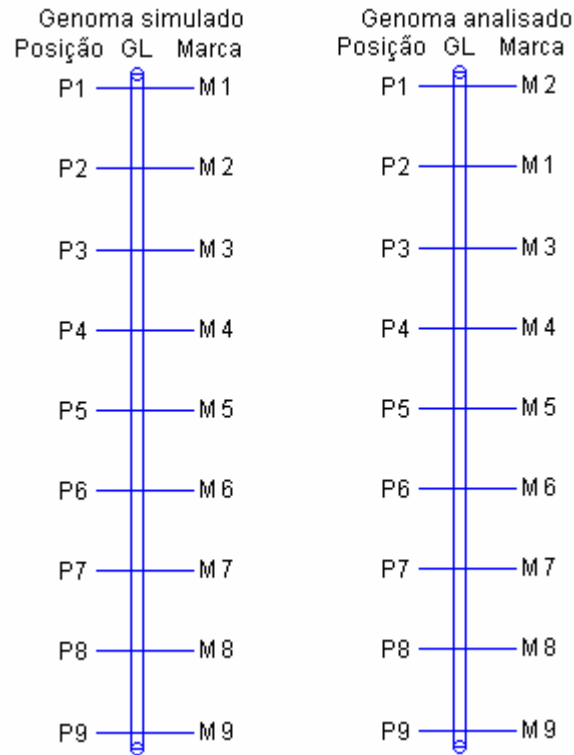


Figura 3 – Grupos de ligação simulado e analisado com nove marcas e inversão das marcas M2 e M1, no grupo de ligação analisado.

Essa correlação não é afetada pelas distâncias entre marcadores. Para estimar os efeitos das mudanças nas distâncias entre os marcadores no genoma analisado em relação ao genoma originalmente simulado, foi utilizado o coeficiente de estresse.

3.3.6. Estresse

O coeficiente de estresse (S) é utilizado como medida de adequação da representação gráfica de medidas de dissimilaridade convertidas em escores relativos às variáveis x e y em estudos de divergência genética (CRUZ e CARNEIRO, 2003). A sua descrição para análise de genomas é demonstrada a seguir:

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

em que S é o valor estimado do estresse, em porcentagem, para o grupo de ligação do genoma analisado; d_{ok} é a distância entre marcas adjacentes m_k e m_{k+1} no grupo de ligação do genoma simulado; e d_k é a distância entre marcas adjacentes m_k e m_{k+1} no grupo de ligação do genoma analisado ($k=1, \dots, m-1$), em que m é o número de marcadores no grupo de ligação do genoma simulado e no grupo de ligação analisado.

Se os marcadores no genoma analisado mantiverem as mesmas distâncias do genoma simulado, o valor estimado do estresse será zero.

3.4. Testes de comparação múltipla

As médias das variáveis como tamanho do grupo de ligação, distância média de marcas adjacentes, variância e estresse em cada grupo de ligação obtido em vários tamanhos de população dentro de cada nível de saturação do genoma foram comparados pelo teste de Tukey a 1% (erro tipo I), com o auxílio do aplicativo computacional GENES (CRUZ, 2004). Também foram comparadas as médias gerais (médias de todos os grupos de ligação) de cada tamanho de população dentro de cada nível de saturação do genoma.

4. RESULTADOS E DISCUSSÃO

4.1. Considerações gerais

Tanto nas populações de retrocruzamento quanto nas populações F_2 não houve a recuperação dos 10 grupos de ligação esperados, para a saturação de cinco marcas por grupo de ligação em amostras com 50 indivíduos, ocorrendo ora a formação de maior ou menor número de grupos de ligações. Com isso, não puderam ser feitas análises devido à falta de informações ou informações errôneas geradas a partir dessas amostras. Exemplos de formação de grupos de ligação em número diferente do sabidamente existente são apresentados na Figura 4.

Observa-se, na Figura 4, a formação de 14 grupos de ligação com apenas duas marcas cada, dois grupos de ligação com três marcas e apenas um grupo de ligação com cinco marcas, que é o número de marcas sabidamente existente em cada grupo de ligação. Assim, pode-se ter uma idéia preliminar do que acontece quando não se têm número de indivíduos e de marcas adequados. Os grupos que, originalmente, comportariam as cinco marcas não são identificados, e no lugar deles são identificados grupos menores, formados por fragmentos de grupos de ligação como é o caso, por exemplo, dos grupos de ligações 1 e 8, que deveriam estar formando um só grupo de ligação. Com base nos dados originais deveriam existir 10 grupos de ligação com cinco marcas cada, totalizando 50 marcas, pelo

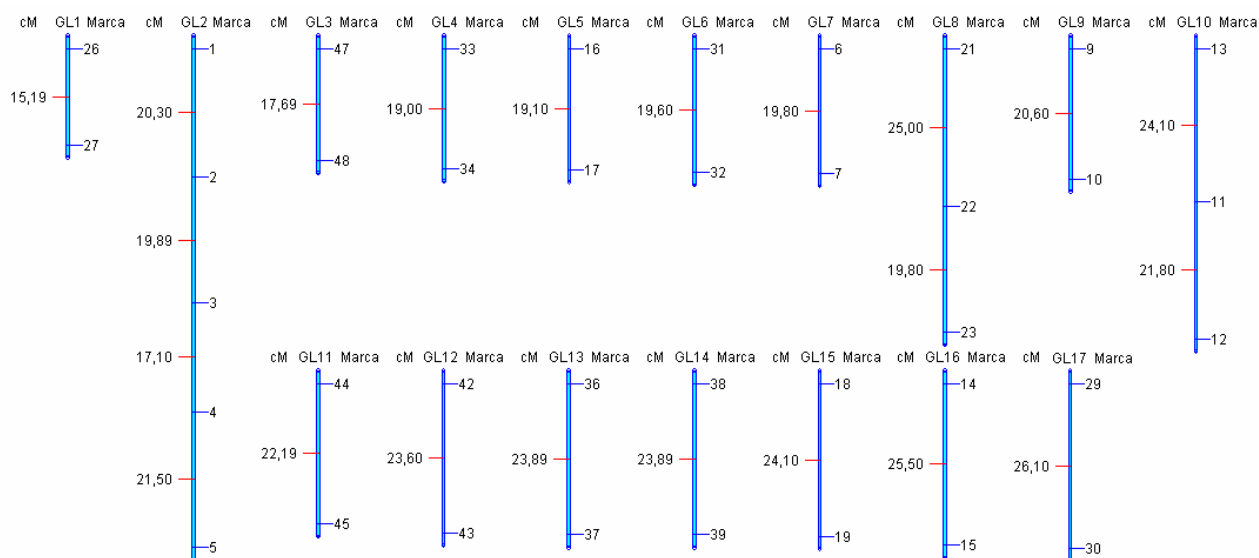


Figura 4 – Representação de uma repetição de população com 50 indivíduos e saturação de cinco marcas por grupo de ligação, um total de sete grupos a mais e 11 marcas a menos que o sabidamente existente. Os grupos deveriam todos conter cinco marcas.

que pode ser visto, apesar dos 17 grupos de ligação existem 39 marcas, 11 marcas a menos, ou seja, com um número inadequado de indivíduos foi gerada uma população com sete grupos de ligação a mais e 11 marcas a menos em relação ao total sabidamente existente.

Na Figura 5 é apresentado um exemplo com amostra de 50 indivíduos e 10 marcas por grupo de ligação da formação. Nesse caso houve a “quebra” do grupo de ligação 6, dando origem ao grupo de ligação 11, formado mais um grupo de ligação.

A inversão também é um fator importante a ser verificado para se ter uma idéia da confiabilidade dos dados obtidos. Na Figura 6 são apresentados exemplos de inversões.

As inversões podem se dar de várias formas, havendo casos em que o grupo de ligação é formado, mas com alterações da ordem de uma ou mais marcas. Além disso, as inversões podem acontecer, também, dentro de grupos de ligação já invertidos (Figura 7), dificultando a decisão de considerar o determinado grupo invertido ou não, para que não ocorram erros durante análises ou formação de matrizes de informação por parte de programas utilizados.

Na Figura 7 são apresentados alguns grupos de ligação que contêm marcas invertidas. No primeiro grupo de ligação, da esquerda para a direita é mostrada uma seqüência de três marcas invertidas e que, além de invertidas, estão deslocados de seu devido lugar. As setas indicam os pares de marcadores invertidos nos grupos de ligação.

Outro fator a desqualificar as populações para análise é a junção de grupos de ligação. Essa junção pode ser total, em que um grupo inteiro se liga a outro grupo inteiro ou parcial, quando um grupo de ligação se liga à parte de outro grupo de ligação.

Na Figura 8 é apresentada uma situação em que um grupo de ligação se ligou a uma parte de outro grupo, não ocorrendo diminuição do número de grupos de ligação. Como pode ser visto, esse fenômeno resultou na formação de três grupos de ligação, sendo um formado por um grupo de ligação inteiro e parte de outro grupo de ligação e os outros dois, apenas fragmentos de grupo de ligação, com quatro marcas cada um.

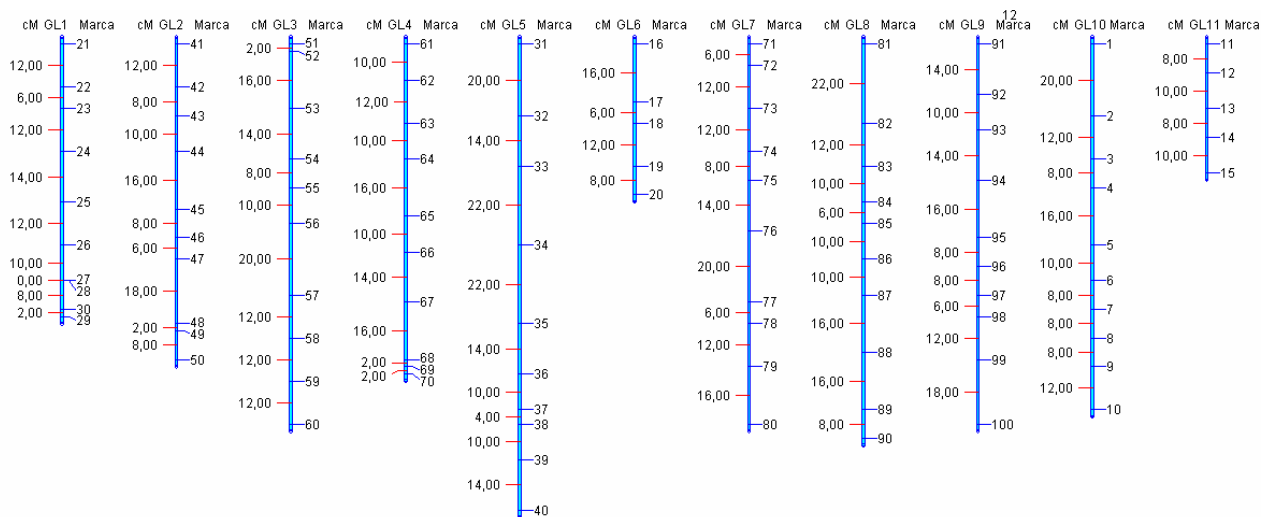


Figura 5 – Representação da formação de 11 grupos de ligação, um a mais que o número sabidamente existente. Os grupos deveriam conter 10 marcas.

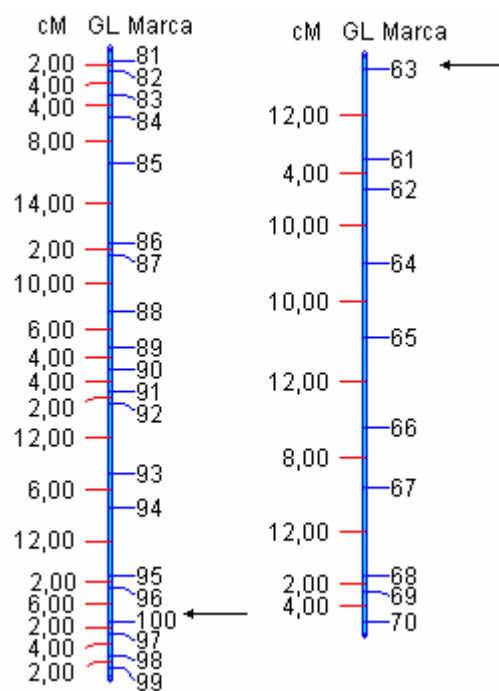


Figura 6 – Representação da posição inadequada de marcas em dois grupos de ligação com saturações diferentes.

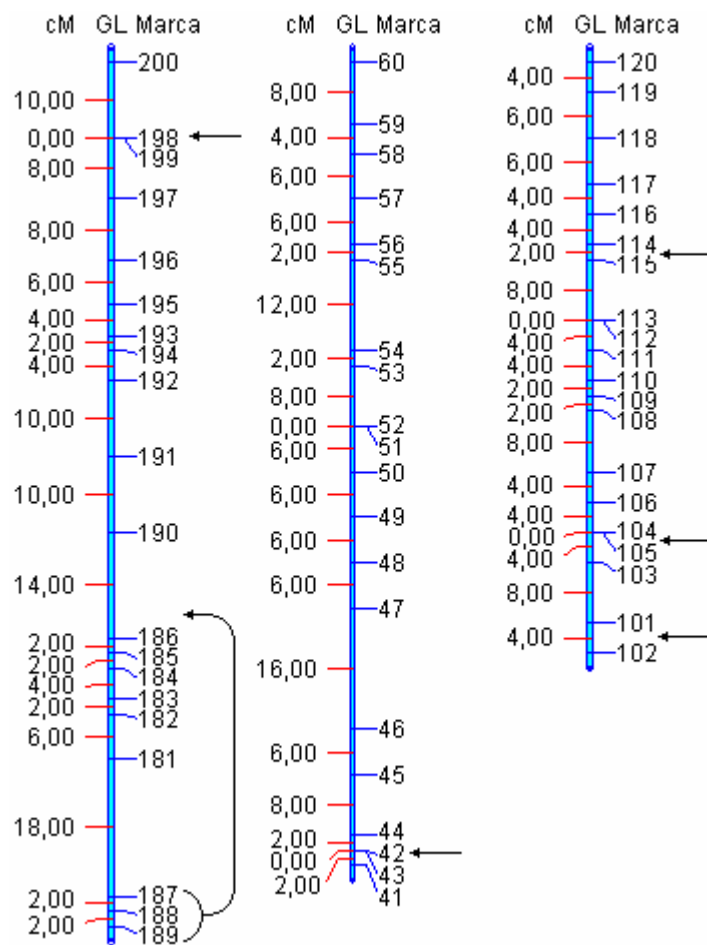


Figura 7 – Representação da inversão de marcas e de grupo de marcas em grupos de ligação já invertidos.

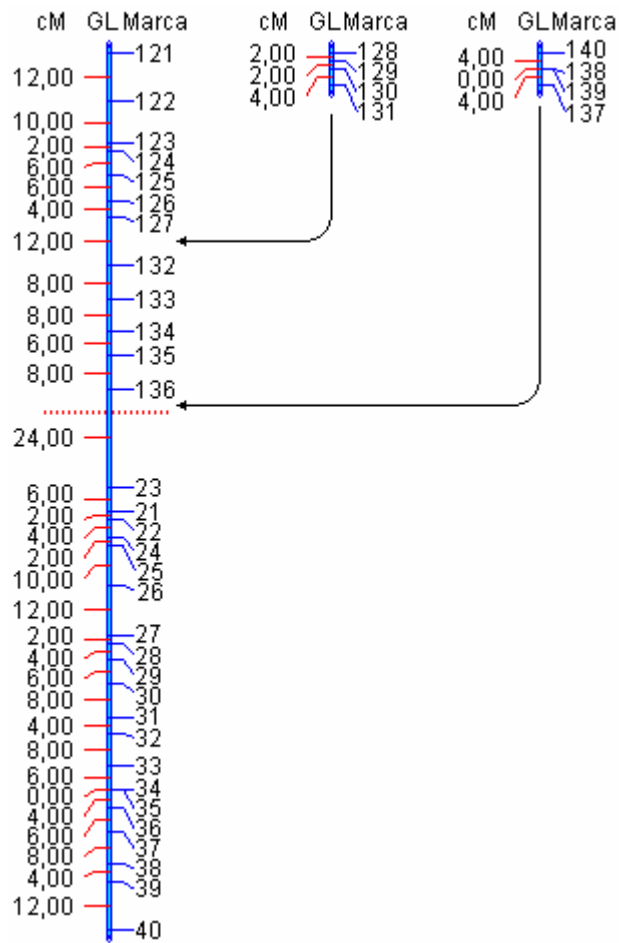


Figura 8 – Representação da ligação entre um grupo de ligação e parte de outro grupo de ligação. O trecho pontilhado separa visualmente um grupo do outro.

Na Figura 9 é mostrada a ligação entre dois grupos de ligação inteiros e dois outros grupos de ligação normais, porém de tamanho reduzido. O menor tamanho dos grupos de ligação provocou a sobreposição de marcas, não permitindo saber ao certo se a seqüência de marcas e a distância entre elas seriam corretas. Sabe-se que são anomalias por conhecimento prévio dos grupos de ligação originais usados para simulação.

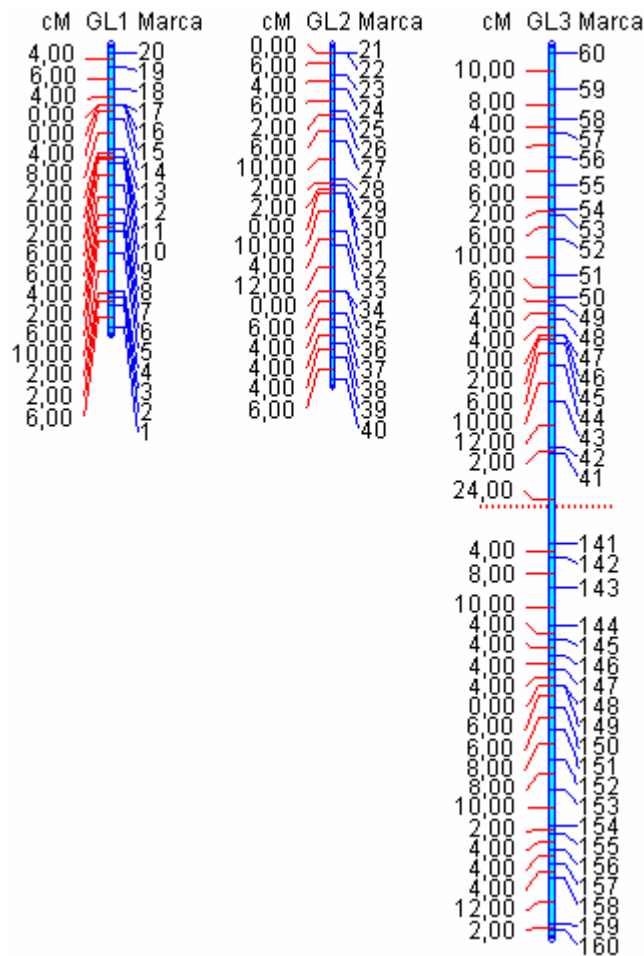


Figura 9 – Representação da ligação entre dois grupos de ligação inteiros. O trecho pontilhado separa visualmente um grupo do outro.

Os defeitos encontrados em populações de retrocruzamento e F_2 em amostras com 50 indivíduos são encontrados em escala muito menor em amostras com maior número de indivíduos e nível de saturação mais elevado e tendem a diminuir à medida que o número de indivíduos da amostra e o número de marcas por grupo de ligação aumentam. Quando o número de indivíduos é muito pequeno, o aumento do número de marcas não é suficiente para que as anomalias não ocorram.

Como dito anteriormente, Ooijen (1992) comentou que o procedimento de mapeamento de QTL de Lander e Botstein (1989) permite determinar, com limitações, a posição de QTLs em um mapa. QTLs com pequeno efeito aditivo

($\hat{\sigma}_{\text{explicada}}^2 = 1\%$) são muito difíceis de serem detectadas, sendo necessária a análise de pelo menos 200 indivíduos, para que efeitos aditivos dessa magnitude sejam detectados. Considerando uma situação prática, 400 indivíduos parecem ser o menor número viável para trabalhos com RFLP, a menos que se esteja interessado apenas em genes de efeito muito grande ($\hat{\sigma}_{\text{explicada}}^2 > 10\%$).

Ooijen (1992) ressaltou que a precisão da estimação da localização de QTLs foi decepcionante, especialmente para o propósito de clonagem de genes. Contudo, para o propósito de melhoramento a precisão parece adequada, dependendo da importância dos outros genes na área do gene a ser mapeado. É claro que genes de efeito genotípico maior que do gene estudado serão mapeados com maior precisão.

4.2. Índice de informatividade e desvio-padrão em populações F₂

A seguir são apresentados valores do índice de informatividade, variância e desvio-padrão das estimativas das quatro frequências de recombinação (0,20; 0,10; 0,0667; e 0,05), nos sete tamanhos de população F₂ simulados neste trabalho.

Ao ser analisado o efeito do tamanho da população, não considerando a estimativa da frequência de recombinação, é verificado que o índice de informatividade será maior quanto maior for o tamanho da população avaliada. Como exemplo, pode-se observar o valor do índice de informatividade relacionado à frequência de recombinação de 0,05, que são de 1.994,76 e 31.916,25, para as populações com 50 e 800 indivíduos, respectivamente.

Como a variância é inversamente relacionada ao índice de informatividade, $V(r) = \frac{1}{I(r)}$, como visto anteriormente, o valor do desvio-padrão diminui à medida que o valor do índice de informatividade aumenta. Ainda considerando frequência de recombinação de 0,05, vê-se que o valor da variância varia de $5,01 \times 10^{-4}$ a $3,13 \times 10^{-5}$ e o valor do desvio-padrão, de 0,0223 e 0,0055, nas populações com 50 e 800 indivíduos, respectivamente.

Analisando-se o efeito da magnitude da estimativa da frequência de recombinação, é observado que, quanto menor a estimativa da frequência de recombinação, maior o valor do índice de informatividade e menores os valores do desvio-padrão, considerado um mesmo tamanho de população. Essas relações são mais bem entendidas quando, por exemplo, dentro de populações com 50 indivíduos, analisadas as estimativas de frequências de recombinação de 0,05; 0,0667; 0,10; e 0,20, em que os valores do índice de informatividade são, respectivamente, 1994,76; 1492,92; 989,15; e 477,94, e os desvios são 0,0223; 0,0258; 0,0317; e 0,0457.

Com base no que pode ser observado no Quadro 4, quanto maior o número de indivíduos F_2 genotipados para a construção de mapas genéticos, maior a confiabilidade dos resultados obtidos e, quanto maior a perspectiva de saturação do mapa por marcadores, maior a confiabilidade das estimativas de frequência de recombinação obtida e dos mapas construídos a partir dessas estimativas.

4.3. Recuperação de grupos de ligação em populações F_2

A princípio, o número de grupos de ligação esperado no processo de mapeamento das populações F_2 simuladas seria de 10 grupos de ligação, que é o número de grupos de ligação que se tem no genoma original usado para a simulação das populações. Apesar disso, o número de grupos de ligação obtidos variou para um maior ou menor número de grupos de ligação de acordo com o nível de saturação do genoma e com o tamanho da população utilizada para geração de cada população simulada.

4.3.1. Saturação dos grupos de ligação

Para a população com 100 indivíduos, 14 repetições recuperaram 10 grupos de ligação, 25 apresentaram 11 grupos de ligação, 21 exibiram 12 grupos de ligação, 10 apresentaram 13 grupos de ligação e cinco apresentaram 14 grupos de ligação. Nessa população houve incidência de marcas não ligadas de 10,97%, em relação ao número de marcas totais na população.

Quadro 4 – Índice de informatividade e desvio-padrão representados para sete tamanhos de população de retrocruzamento, em quatro níveis de saturação

Frequência Recombinação	T.P.	Índice de Informatividade	Desvio-Padrão
0,20 (5 marcas)	50	477,94	0,0457
	100	955,88	0,0323
	150	1433,82	0,0264
	200	1911,76	0,0228
	400	3823,52	0,0161
	600	5735,29	0,0132
	800	7647,05	0,0114
0,10 (10 marcas)	50	989,15	0,0317
	100	1978,31	0,0224
	150	2967,47	0,0183
	200	3956,63	0,0158
	400	7913,27	0,0112
	600	11869,91	0,0091
	800	15826,55	0,0079
~0,0667(15 marcas)	50	1492,92	0,0258
	100	2985,85	0,0183
	150	4478,78	0,0149
	200	5971,71	0,0129
	400	11943,43	0,0091
	600	17915,15	0,0074
	800	23886,87	0,0064
0,05 (20 marcas)	50	1994,76	0,0223
	100	3989,53	0,0158
	150	5984,29	0,0129
	200	7979,06	0,0111
	400	15958,12	0,0079
	600	23937,19	0,0064
	800	31916,25	0,0055

T.P. = tamanho da população.

Para a população com 150 indivíduos, 32 repetições recuperaram 10 grupos de ligação, 27 apresentaram 11 grupos de ligação, 23 apresentaram 12 grupos de ligação e 6 apresentaram 13 grupos de ligação. Em relação ao número de marcas não ligadas, houve redução quanto à população com 100 indivíduos, e a porcentagem passou de 10,97 % para 6,92 %.

Na Figura 10 estão representadas as populações com 50, 100, 150, 200, 400, 600 e 800 indivíduos, com saturação de cinco marcas por grupo de ligação. Como pode ser observado nessa figura, em nenhuma das populações foi recuperado o conjunto inicial de 10 grupos de ligação em todas as repetições simuladas, sendo a população com 800 indivíduos a que mais se aproximou desse objetivo, recuperando os 10 grupos em 99 das 100 repetições.

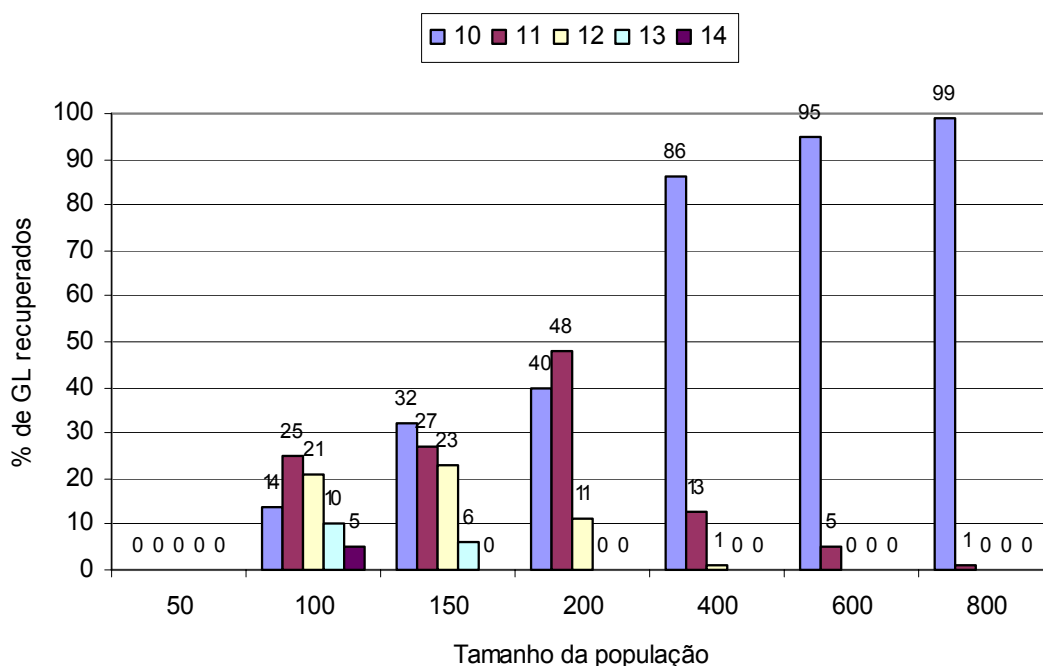


Figura 10 – Da esquerda para a direita, número de 10, 11, 12, 13 e 14 grupos de ligação obtidos para saturações de cinco marcas em populações F_2 com 50, 100, 150, 200, 400, 600 e 800 indivíduos.

Para a população com 200 indivíduos, 40 repetições recuperaram 10 grupos de ligação, 48 apresentaram 11 grupos de ligação e 11 tiveram 12 grupos de ligação. Nessa população houve uma incidência de 4,30% de marcas não-ligadas.

Para a população com 400 indivíduos, 86 repetições recuperaram 10 grupos de ligação, 13 apresentaram 11 grupos de ligação e uma exibiu 12 grupos, havendo a presença de marcas não ligadas de 0,88%. Na população com 600 indivíduos, 95 repetições recuperaram 10 grupos de ligação e cinco exibiram 11 grupos de ligação, havendo 0,34% de marcas não-ligadas. Finalmente, na população com 800 indivíduos, 99 repetições recuperaram 10 grupos de ligação e uma exibiu 11 grupos com 0,06% de marcas não-ligadas.

Foi constatada a não-recuperação de grupos de ligação em 25 repetições em população com 100 indivíduos, 12 repetições em população com 150

indivíduos e uma repetição na população com 200 indivíduos. Nas demais houve formação de grupos de ligação em todas as repetições.

É interessante notar a tendência de aumento na obtenção do número original de grupos de ligação à medida que cresce o número de indivíduos nas populações. A diminuição no número de marcas não-ligadas parece ser inversamente relacionada ao tamanho da população, ou seja, com o aumento no número de indivíduos, vê-se a redução no número de marcas não-ligadas (Quadro 5).

Pela saturação de cinco marcas por grupo de ligação, pode ser observado que a quantidade de marcas não-ligadas tende a diminuir com o aumento no número de indivíduos da população. Isso pode ser visto claramente quando se compara a quantidade de marcas não-ligadas na população com 100 indivíduos com a população com 800 indivíduos, 428 ou 10,97% contra três ou 0,06%. Analisando apenas o tamanho da população, pode-se inferir que, quanto maior o número de indivíduos dentro de uma população, mais confiáveis os dados obtidos.

Um fato interessante é apresentado no Quadro 5, ou seja, a identificação de anomalia em relação aos grupos de ligação nas populações com 100 indivíduos nas saturações de 10 e 15 marcas por grupo de ligação. Nessas populações existe a formação de um grupo de ligação menor, com ausência de algumas marcas, mas que não são não-ligadas e sim marcas que estão ligadas a outro grupo de ligação, que está completo em relação às marcas. Desse modo, entre os grupos de ligações normais aparece um com número de marcas maior do que a saturação da população e outro grupo com menor número de marcas e nenhuma marca não-ligada. Apesar de acontecer em frequência muito pequena, é um dado importante a ser relatado.

No Quadro 5 é apresentada a porcentagem de marcas não-ligadas, obtidas em todas as repetições que apresentaram a recuperação de grupos de ligação, independentemente da quantidade de grupos formados.

Quadro 5 – Saturação e porcentagem de marcas não-ligadas em relação ao número de repetições avaliadas em sete tamanhos de população F₂

Saturação	T.P.	MNL	Rep	%MNL
5 marcas	50	-	-	-
	100	428	78	10,97
	150	291	84	6,92
	200	209	97	4,30
	400	44	100	0,88
	600	17	100	0,34
	800	3	100	0,06
10 marcas	50	27	84	0,32
	100	3 e -3	100	*
	150	0	100	0
	200	0	100	0
	400	0	100	0
	600	0	100	0
	800	0	100	0
15 marcas	50	0	88	0
	100	9 e -9	100	*
	150	0	100	0
	200	0	100	0
	400	0	100	0
	600	0	100	0
	800	0	100	0
20 marcas	50	5	92	0,02
	100	0	98	0
	150	0	100	0
	200	0	100	0
	400	0	100	0
	600	0	100	0
	800	0	100	0

[†] indica que a primeira linha está vazia por falta de informações necessárias para geração de grupos de ligação com 50 indivíduos e concentração de cinco marcas por grupo de ligação; saturação = número de marcas por grupo de ligação, do genoma simulado; T.P. = tamanho da população; MNL = número de marcas não-ligadas; Rep = número de repetições com 10 grupos de ligação; % MNL = porcentagem de marcas não-ligadas em relação ao número de repetições avaliadas; e * = anomalia, representação de grupo de ligação em que parte foi ligada a outro grupo de ligação.

Considerando, agora, a saturação de 10 marcas por grupo de ligação e populações com 50, 100, 150, 200, 400, 600 e 800 indivíduos em relação à recuperação de grupos de ligação e número de marcas não-ligadas, vê-se que (Figura 11) houve uma recuperação de 10 grupos de ligação em 76 repetições, 11 grupos de ligação em sete repetições e uma repetição com nove grupos de ligação, na população com 50 indivíduos, chegando a um total de 84 repetições

com recuperação de grupos de ligação; nessa população, a presença de marcas não-ligadas foi de 0,32% (Quadro 5). Nas populações com 100, 150, 200, 400, 600 e 800 indivíduos foram observadas a recuperação de 10 grupos de ligação em todas as repetições e a ausência de marcas não-ligadas.

Na saturação de 15 marcas por grupo de ligação, na população com 50 indivíduos houve a recuperação de 88 repetições com 10 grupos de ligação (Figura 11). As demais populações recuperaram 10 grupos de ligação em todas as repetições. Também nesta saturação não houve a constatação de marcas não-ligadas em nenhuma das populações analisadas.

Na saturação de 20 marcas por grupo de ligação na população com 50 indivíduos houve a recuperação de 10 grupos de ligação em 92 repetições e 11 grupos de ligação em duas repetições. Em relação ao número de marcas não-ligadas, pode-se observar, no Quadro 5, que, apesar da maior saturação nesse caso, houve a identificação de 0,02% de marcas não-ligadas, enquanto na saturação de 15 marcas por grupo de ligação não ocorreu a identificação de marcas não-ligadas. Isso vem confirmar a idéia de que, mesmo em um alto nível de saturação, a avaliação de reduzido número de indivíduos não gera dados de boa confiabilidade.

Nesse caso também houve a recuperação de 10 grupos de ligação em todas as repetições das demais populações e nenhuma identificação de marcas não-ligadas.

O nível de saturação com cinco marcas e população com 50 indivíduos foi desconsiderado por não ter obtido grupos de ligação em suas repetições (Figura 10).

Na Figura 11 é apresentada a porcentagem de grupos de ligação recuperados a partir da simulação feita utilizando-se saturações de 10, 15 e 20 marcas por grupo de ligação, em populações com 50 indivíduos.

Semelhantemente ao que se observa na Figura 10, à medida que o número de marcas por grupo de ligação aumenta, o número de repetições que recuperam 10 grupos de ligação também se eleva, e, mesmo aumentando a saturação, não acontece recuperação de grupos de ligação em 100% das repetições, devido ao baixo número de indivíduos nas populações.

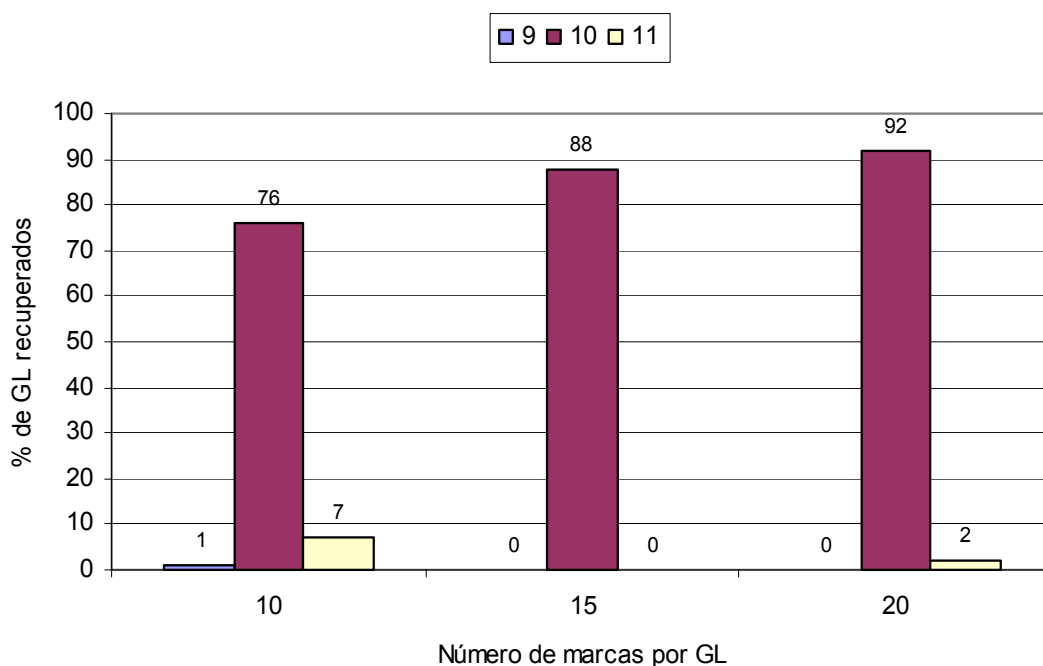


Figura 11 – Da esquerda para a direita, número de 9, 10 e 11 grupos de ligação obtidos nas saturações de 10, 15 e 20 marcas em populações F_2 com 50 indivíduos.

Considerando tanto a Figura 10 quanto a Figura 11 e aumentando o número de indivíduos ou a saturação independentemente uma da outra, observa-se que existe diminuição no número de repetições sem grupos de ligação recuperados, bem como aumento no número de repetições com recuperação do número original de grupos de ligação.

Segundo Weller (1986) e Patterson et al. (1988), uma porção significativa dos efeitos de características quantitativas é de uma magnitude que poderia ser prontamente detectada em experimentos com aproximadamente 1.000 indivíduos de população F_2 ou de retrocruzamento. Além disso, se duas linhagens diferem em muitos marcadores localizados ao longo do genoma, uma única geração F_2 permitirá que sejam feitas avaliação total e mapeamento do genoma inteiro de todas as características quantitativas avaliadas. Porém, a quantidade de informação produzida por esse tipo de análise é crítica, dependendo do número de marcadores polimórficos que diferem nas linhas parentais.

4.3.2. Correlação de Spearman entre medidas de distância

A correlação de Spearman foi utilizada para a identificação de inversão de posição de marcas dentro de cada um dos 10 grupos de ligação formados no mapeamento. A avaliação da inversão só foi feita em repetições em que foram recuperados os 10 grupos de ligação.

Por se tratar de populações hipotéticas, foi considerado não ser necessária a especificação das repetições e em quais grupos de ligação ocorreu inversão, sendo relatado apenas o número de repetições a recuperar 10 grupos de ligação e quantos desses grupos apresentaram inversões, independentemente do número de inversões ocorridas dentro de cada repetição.

Obtendo valores de correlação de Spearman iguais a 1, entende-se que as ordens das marcas nos grupos de ligação obtidos no mapeamento das populações segregantes não foram alteradas em relação à ordem previamente conhecida do genoma original, utilizado para a geração das simulações. Porém, caso os valores de correlação de Spearman fossem menores que 1, a indicação seria de que a ordem das marcas nos grupos de ligação obtidos no mapeamento da população segregante fosse alterada em relação à ordem do genoma de referência.

Repetições com presença de valores de correlação de Spearman menores que 1, em um ou mais dos 10 grupos de ligação, foram obtidas nas populações segregantes, nos quatro níveis de saturação de genoma estudados (Quadro 6). No nível de saturação com cinco marcas, das 14 repetições com 10 grupos de ligação a população de 100 indivíduos apresentou apenas uma repetição com valor de correlação de Spearman diferente de 1; nas demais populações, nesse grau de saturação não se identificaram mais inversões. Nas outras saturações houve a recuperação de grupos de ligação na população com 50 indivíduos. Na saturação de 10 marcas, 76 recuperaram 10 grupos de ligação e 18 apresentaram inversão, enquanto na população com 100 indivíduos 100 recuperaram 10 grupos de ligação e apenas um apresentou inversão; na saturação de 15 marcas, na população com 50 indivíduos 89 recuperaram 10 grupos de ligação e 48 apresentaram inversão; na população com 100 indivíduos, 100 recuperaram 10

grupos de ligação e apenas um apresentou inversão; na saturação de 20 marcas, 91 recuperaram 10 grupos de ligação e 85 exibiram inversão; na população com 100 indivíduos, 98 recuperaram 10 grupos de ligação e três apresentaram inversão; e, na população com 150 indivíduos, 100 recuperaram 10 grupos de ligação e um teve inversão.

No Quadro 6 é apresentado o número de inversões identificadas pela correlação de Spearman, em repetições que recuperaram 10 grupos de ligação, nos sete tamanhos de população F_2 e quatro níveis de saturação utilizados.

Quadro 6 – Nível de saturação em número de marcas por grupo de ligação, tamanho da população, número de repetições com 10 grupos de ligação e número de inversões detectadas em populações F_2

Saturação	T.P.	Rep	NI
5 marcas	50	-	- [†]
	100	14	1
	150	32	0
	200	40	0
	400	86	0
	600	95	0
	800	99	0
10 marcas	50	76	18
	100	100	1
	150	100	0
	200	100	0
	400	100	0
	600	100	0
	800	100	0
15 marcas	50	89	48
	100	100	1
	150	100	0
	200	100	0
	400	100	0
	600	100	0
	800	100	0
20 marcas	50	91	85
	100	98	3
	150	100	1
	200	100	0
	400	100	0
	600	100	0
	800	100	0

[†] indica que a primeira linha está vazia por falta de informações necessárias para geração de grupos de ligação com 50 indivíduos e concentração de cinco marcas por grupo de ligação; saturação = número de marcas por grupo de ligação, do genoma simulado; T.P. = tamanho da população; Rep = número de repetições com 10 grupos de ligação; e NI = número de inversões detectadas.

Em todos os outros tamanhos de população nos diferentes níveis de saturação, todos recuperaram 10 grupos de ligação em 100% das repetições e não houve inversões.

É interessante observar, no Quadro 6, que com o aumento da saturação, em populações com 50 indivíduos, houve aumento também no número de repetições que recuperaram 10 grupos de ligação, porém ainda ocorreu elevação no número de inversões. Na população com 100 indivíduos, considerando as saturações de 10, 15 e 20 marcas, cresceu o número de inversão na população com maior grau de saturação. Também ocorreu inversão na população com 150 indivíduos, na maior saturação. A causa desse incremento no número de inversões, apesar da elevação do número de marcas é, por ora, desconhecido, sendo que a partir de populações com 200 indivíduos, independentemente do nível de saturação utilizado, não ocorreram inversões.

4.3.3. Tamanho dos grupos de ligação

Com relação ao tamanho médio de cada grupo de ligação, foi conhecida a média aritmética dos tamanhos dos grupos de ligação obtidos nas repetições com 10 grupos de ligação. Os valores dos tamanhos médios são apresentados no Quadro 7.

O tamanho esperado nos grupos de ligação após o mapeamento das populações segregantes era de 100 cM, uma vez que esse é o tamanho de cada grupo de ligação nos quatro níveis de saturação do genoma utilizado para simulação das populações.

Como explicado anteriormente, na avaliação das médias dos tamanhos dos grupos de ligação foi feita a análise de variância seguida por teste comparativo de médias (teste de Tukey), em que as diferenças entre as médias foram avaliadas com nível de significância de 0,01. O teste de Tukey permitiu a avaliação estatística do efeito do tamanho da população no tamanho de grupos de ligação formados no mapeamento das populações simuladas.

Quadro 7 – Média aritmética do tamanho de 10 grupos de ligação em sete tamanhos de populações (T.P.) F₂, em quatro níveis de saturação

Saturação	T.P.	Grupos de Ligação										Média Geral													
		1	2	3	4	5	6	7	8	9	10														
5 marcas	50	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-			
	100	87,56 (b)	91,75 (a)	85,11 (b)	88,28 (b)	96,97 (a)	88,56 (b)	88,30 (b)	92,92 (a)	94,39 (a)	93,17 (a)	90,70 (c)	90,70 (c)	95,37 (b)	97,88 (b)	98,80 (a)	98,71 (a)	98,80 (a)	98,80 (a)	98,80 (a)	98,80 (a)	98,80 (a)	98,80 (a)	98,80 (a)	
	150	92,71 (b)	92,93 (a)	97,11 (ab)	93,86 (ab)	94,55 (a)	96,45 (ab)	94,18 (ab)	95,81 (a)	97,27 (a)	98,67 (a)	98,96 (a)	98,71 (a)	97,88 (b)	98,80 (a)	98,80 (a)	98,80 (a)	98,80 (a)	98,80 (a)	98,80 (a)	98,80 (a)	98,80 (a)	98,80 (a)	98,80 (a)	98,80 (a)
	200	96,59 (ab)	98,12 (a)	96,72 (ab)	98,40 (ab)	97,11 (a)	96,74 (ab)	98,78 (ab)	98,67 (a)	98,96 (a)	98,96 (a)	98,71 (a)	97,88 (b)	98,80 (a)	98,80 (a)	98,80 (a)	98,80 (a)	98,80 (a)	98,80 (a)	98,80 (a)	98,80 (a)	98,80 (a)	98,80 (a)	98,80 (a)	98,80 (a)
	400	101,05 (a)	99,98 (a)	99,86 (a)	100,08 (ab)	101,40 (a)	101,12 (a)	101,13 (a)	100,58 (a)	101,31 (a)	101,31 (a)	101,04 (a)	100,75 (a)	100,91 (a)	100,91 (a)	100,91 (a)	100,91 (a)	100,91 (a)	100,91 (a)	100,91 (a)	100,91 (a)	100,91 (a)	100,91 (a)	100,91 (a)	100,91 (a)
	800	100,78 (a)	101,13 (a)	101,41 (a)	101,17 (a)	101,28 (a)	101,22 (a)	101,31 (a)	101,31 (a)	101,31 (a)	101,31 (a)	101,31 (a)	101,31 (a)	101,31 (a)	101,31 (a)	101,31 (a)	101,31 (a)	101,31 (a)	101,31 (a)	101,31 (a)	101,31 (a)	101,31 (a)	101,31 (a)	101,31 (a)	101,31 (a)
10 marcas	50	102,86	104,03	102,57	100,28	101,91	105,32	103,09	102,25	102,91	103,52	102,87 (a)	102,87 (a)	102,87 (a)	102,87 (a)	102,87 (a)	102,87 (a)	102,87 (a)	102,87 (a)	102,87 (a)	102,87 (a)	102,87 (a)	102,87 (a)	102,87 (a)	
	100	103,5	103,78	102,83	102,60	102,53	102,27	103,53	102,58	102,50	103,96	103,01 (a)	103,01 (a)	103,01 (a)	103,01 (a)	103,01 (a)	103,01 (a)	103,01 (a)	103,01 (a)	103,01 (a)	103,01 (a)	103,01 (a)	103,01 (a)	103,01 (a)	103,01 (a)
	150	103,06	103,56	103,57	102,21	102,60	102,14	102,07	103,22	102,45	102,26	102,71 (a)	102,71 (a)	102,71 (a)	102,71 (a)	102,71 (a)	102,71 (a)	102,71 (a)	102,71 (a)	102,71 (a)	102,71 (a)	102,71 (a)	102,71 (a)	102,71 (a)	102,71 (a)
	200	102,43	103,45	101,96	102,78	103,32	102,93	102,85	102,57	103,49	103,04	102,88 (a)	102,88 (a)	102,88 (a)	102,88 (a)	102,88 (a)	102,88 (a)	102,88 (a)	102,88 (a)	102,88 (a)	102,88 (a)	102,88 (a)	102,88 (a)	102,88 (a)	102,88 (a)
	400	102,73	102,41	102,71	102,76	102,55	102,88	102,88	102,9	102,98	102,84	102,76 (a)	102,76 (a)	102,76 (a)	102,76 (a)	102,76 (a)	102,76 (a)	102,76 (a)	102,76 (a)	102,76 (a)	102,76 (a)	102,76 (a)	102,76 (a)	102,76 (a)	102,76 (a)
	800	103,01	102,40	102,98	102,70	103,01	102,74	102,60	102,65	103,00	103,11	102,82 (a)	102,82 (a)	102,82 (a)	102,82 (a)	102,82 (a)	102,82 (a)	102,82 (a)	102,82 (a)	102,82 (a)	102,82 (a)	102,82 (a)	102,82 (a)	102,82 (a)	102,82 (a)
15 marcas	50	104,04	100,04	101,60	101,83	101,21	104,35	103,68	102,30	102,80	103,78	102,56 (a)	102,56 (a)	102,56 (a)	102,56 (a)	102,56 (a)	102,56 (a)	102,56 (a)	102,56 (a)	102,56 (a)	102,56 (a)	102,56 (a)	102,56 (a)	102,56 (a)	
	100	102,90	102,81	101,18	104,03	101,53	102,79	102,33	102,64	103,20	103,66	102,71 (a)	102,71 (a)	102,71 (a)	102,71 (a)	102,71 (a)	102,71 (a)	102,71 (a)	102,71 (a)	102,71 (a)	102,71 (a)	102,71 (a)	102,71 (a)	102,71 (a)	
	150	102,24	101,67	102,63	102,13	102,18	102,08	103,01	102,81	101,95	101,69	102,24 (a)	102,24 (a)	102,24 (a)	102,24 (a)	102,24 (a)	102,24 (a)	102,24 (a)	102,24 (a)	102,24 (a)	102,24 (a)	102,24 (a)	102,24 (a)	102,24 (a)	102,24 (a)
	200	101,57	102,66	102,77	102,34	102,93	102,17	102,49	102,18	102,59	102,78	102,45 (a)	102,45 (a)	102,45 (a)	102,45 (a)	102,45 (a)	102,45 (a)	102,45 (a)	102,45 (a)	102,45 (a)	102,45 (a)	102,45 (a)	102,45 (a)	102,45 (a)	102,45 (a)
	400	102,94	102,48	102,19	102,60	103,00	102,49	102,48	102,37	103,17	102,51	102,62 (a)	102,62 (a)	102,62 (a)	102,62 (a)	102,62 (a)	102,62 (a)	102,62 (a)	102,62 (a)	102,62 (a)	102,62 (a)	102,62 (a)	102,62 (a)	102,62 (a)	102,62 (a)
	800	103,00	102,45	102,52	102,41	102,26	102,25	102,14	102,58	102,15	102,38	102,41 (a)	102,41 (a)	102,41 (a)	102,41 (a)	102,41 (a)	102,41 (a)	102,41 (a)	102,41 (a)	102,41 (a)	102,41 (a)	102,41 (a)	102,41 (a)	102,41 (a)	102,41 (a)
20 marcas	50	102,31	100,12	104,50	101,90	102,74	101,91	103,61	103,06	103,50	104,03	102,77 (a)	102,77 (a)	102,77 (a)	102,77 (a)	102,77 (a)	102,77 (a)	102,77 (a)	102,77 (a)	102,77 (a)	102,77 (a)	102,77 (a)	102,77 (a)	102,77 (a)	
	100	100,55	101,29	101,46	100,85	102,11	103,06	101,39	102,23	100,34	102,70	101,60 (a)	101,60 (a)	101,60 (a)	101,60 (a)	101,60 (a)	101,60 (a)	101,60 (a)	101,60 (a)	101,60 (a)	101,60 (a)	101,60 (a)	101,60 (a)	101,60 (a)	
	150	101,34	101,45	102,37	102,44	102,24	101,64	101,11	101,26	101,69	102,64	101,82 (a)	101,82 (a)	101,82 (a)	101,82 (a)	101,82 (a)	101,82 (a)	101,82 (a)	101,82 (a)	101,82 (a)	101,82 (a)	101,82 (a)	101,82 (a)	101,82 (a)	
	200	101,98	101,82	101,33	101,46	101,78	101,06	101,80	101,96	102,48	102,21	101,79 (a)	101,79 (a)	101,79 (a)	101,79 (a)	101,79 (a)	101,79 (a)	101,79 (a)	101,79 (a)	101,79 (a)	101,79 (a)	101,79 (a)	101,79 (a)	101,79 (a)	
	400	102,15	101,59	102,22	101,38	102,41	102,29	101,57	101,68	101,36	102,04	101,87 (a)	101,87 (a)	101,87 (a)	101,87 (a)	101,87 (a)	101,87 (a)	101,87 (a)	101,87 (a)	101,87 (a)	101,87 (a)	101,87 (a)	101,87 (a)	101,87 (a)	
	800	101,49	101,90	102,15	101,62	101,61	101,88	102,09	101,90	102,17	101,64	101,85 (a)	101,85 (a)	101,85 (a)	101,85 (a)	101,85 (a)	101,85 (a)	101,85 (a)	101,85 (a)	101,85 (a)	101,85 (a)	101,85 (a)	101,85 (a)	101,85 (a)	

¹ indica que a primeira linha está vazia por falta de informações necessárias para a geração de grupos de ligação com 50 indivíduos e concentração de cinco marcas por grupo de ligação; ² indica que dentro dos parêntesis estão as letras correspondentes ao resultado da análise de médias feita pelo teste de Tukey a 1%. Nas colunas, médias seguidas de uma mesma letra não diferem entre si, pelo teste de Tukey.

Verificou-se que, a não ser na saturação de cinco marcas por grupo de ligação, não houve diferença entre as médias dos diferentes tamanhos de população dentro de cada nível de saturação.

No Quadro 7 são apresentados os resultados do teste de Tukey com nível de significância de 0,01. No nível de saturação com cinco marcas por grupo de ligação houve diferença significativa entre os tamanhos dos grupos de ligação das populações de tamanhos diferentes. Como pode ser visto na última coluna do Quadro 7, não houve diferença entre as médias gerais das populações com 400, 600 e 800 indivíduos. As populações com 200 e 150 indivíduos também não diferiram entre si, e a população com 100 indivíduos foi diferente de todas as outras.

Só foi apresentado o resultado do teste de Tukey na última coluna, nos níveis de saturação com 10, 15 e 20 marcas, por ele representar exatamente os mesmos resultados dos 10 grupos de ligação de cada nível de saturação.

Em resumo, pelo Quadro 7 tem-se que, no nível de saturação de cinco marcas por grupo de ligação, à medida que o tamanho da população aumenta, o tamanho médio dos grupos de ligação também aumenta, aproximando do valor originalmente adotado. Nos demais níveis de saturação, não houve variação significativa no tamanho médio dos grupos de ligação com o aumento no tamanho da população.

Uma das maneiras adicionais de observar o comportamento dos tamanhos médios dos grupos de ligação se dá através da análise do desvio-padrão. Com o aumento no número de indivíduos em um mesmo nível de saturação do genoma, pode ser observado que, apesar de o número de marcas por grupo de ligação continuar o mesmo à medida que o número de indivíduos das populações aumenta, o desvio-padrão diminui, o que pode ser verificado no Quadro 8, observando-se os valores de desvio-padrão na última coluna, em todos os níveis de saturação. No nível de saturação de cinco marcas, o desvio passa de 14,45 na população com 100 indivíduos para 2,72 naquela com 800. Considerando os outros níveis de saturação, o desvio-padrão varia de 10,29 a 2,63 em saturação

com 10 marcas, 10,46 a 2,55 na saturação com 15 marcas e 15,03 a 2,52 em saturação com 20 marcas.

Essa tendência de redução na amplitude de variação observada nas médias do tamanho dos grupos de ligação acontece também entre os 10 grupos de ligação dos sete tamanhos de população, a não ser nos grupos de ligação 5 e 10 na saturação de 10 marcas, 4 e 9 na de 15 marcas e 3 e 7 naquela saturação de 20 marcas, em que estão os valores realçados nos respectivos grupos de ligação. Ao contrário do que ocorreu nas demais, no nível de saturação com cinco marcas apenas três grupos de ligação não apresentaram valores que contrariassem a tendência da média, de diminuição na amplitude de variação do tamanho dos grupos de ligação.

É importante observar também que, considerando os três maiores níveis de saturação, de acordo com o representado na última coluna a média dos desvios-padrão é menor que 5 nas populações com 200 ou mais indivíduos e fica abaixo de 6 nos 10 grupos de ligação.

Como visto no Quadro 8, o tamanho dos desvios-padrão diminui com o aumento no tamanho das populações. Por exemplo, no grupo de ligação 1 tem-se que o desvio-padrão foi de 15,51 e 2,48 cM, nos tamanhos de população de 100 e de 800 indivíduos, respectivamente, no menor nível de saturação.

Com base no que é observado nos Quadros 7 e 8, pode-se concluir que, com o aumento no número de indivíduos das populações em três dos quatro níveis de saturação, não existe diferença significativa entre as médias gerais de tamanho dos grupos de ligação, porém há diminuição na amplitude do desvio-padrão com o aumento no número de indivíduos nas populações. No menor nível de saturação, com o aumento do número de indivíduos nas populações o tamanho médio dos grupos de ligação também aumenta, aproximando-se do tamanho originalmente estabelecido; com isso, nesse nível de saturação o desvio-padrão também diminui com o incremento do número de indivíduos das populações.

Quadro 8 – Desvio-padrão do tamanho de 10 grupos de ligação em sete tamanhos de populações (T.P.) F_2 , em quatro níveis de saturação

Saturação	T.P.	Grupos de Ligação										Média Geral	
		1	2	3	4	5	6	7	8	9	10		
5 marcas	50	-	-	-	-	-	-	-	-	-	-	-	¹
	100	15,51	12,35	16,19	17,25	10,93	18,19	18,59	11,88	10,25	13,40	14,45	
	150	12,71	14,12	11,48	12,76	12,44	9,49	12,74	12,2	8,53	7,51	11,39	
	200	9,55	7,66	10,05	6,82	9,34	7,54	9,45	6,62	8,84	8,44	8,43	
	400	3,21	5,52	3,14	5,17	3,00	4,28	5,40	4,69	3,37	3,40	4,11	
	600	2,92	3,99	3,33	4,27	3,69	4,16	3,91	4,00	2,94	4,14	3,73	
	800	2,48	2,75	2,84	2,63	2,71	2,59	2,56	2,2	4,03	2,46	2,72	
10 marcas	50	10,02	9,08	10,10	11,14	11,42	9,43	9,60	11,46	9,91	10,75	10,29	
	100	6,65	8,30	7,74	7,11	6,70	7,47	7,25	8,87	7,05	6,61	7,37	
	150	6,37	6,04	5,86	6,22	5,15	6,22	5,78	6,12	5,51	6,00	5,92	
	200	4,66	5,26	4,62	4,77	5,05	4,72	5,13	5,00	5,05	5,07	4,93	
	400	3,67	3,28	3,49	3,02	3,24	3,42	3,31	3,62	3,41	3,06	3,35	
	600	2,68	2,98	2,84	2,79	2,49	2,88	2,95	2,96	2,47	2,68	2,77	
	800	2,60	2,94	2,41	2,69	2,64	2,70	2,77	2,55	2,33	2,68	2,63	
15 marcas	50	11,31	9,84	10,66	8,97	10,39	9,80	12,03	10,62	11,30	9,76	10,46	
	100	10,99	10,47	7,26	7,04	7,61	8,93	7,90	9,55	11,51	9,28	9,05	
	150	7,17	6,11	6,31	5,22	6,79	6,33	6,39	5,70	6,36	5,94	6,23	
	200	4,38	4,91	5,23	5,94	4,86	4,58	4,83	5,25	5,04	4,80	4,98	
	400	3,66	3,70	3,48	3,46	4,04	3,63	3,73	3,98	3,65	3,69	3,70	
	600	2,70	3,08	3,22	3,02	2,72	2,84	2,94	3,13	2,76	3,16	2,95	
	800	2,45	2,49	2,57	2,57	2,48	2,42	2,58	2,55	2,59	2,82	2,55	
20 marcas	50	13,79	15,07	15,58	13,78	16,58	16,67	13,71	14,02	13,48	17,71	15,03	
	100	6,88	6,35	6,73	7,34	7,47	7,59	7,66	7,02	7,57	7,30	7,19	
	150	4,91	6,14	6,75	7,00	5,96	6,33	5,99	5,58	5,65	6,12	6,04	
	200	4,52	5,29	5,03	4,51	4,87	5,01	4,97	5,52	5,26	4,61	4,95	
	400	4,20	3,54	3,57	3,40	3,64	3,48	3,36	3,87	3,73	3,85	3,66	
	600	3,24	3,18	2,88	2,55	3,01	3,08	3,37	2,91	3,20	3,13	3,05	
	800	2,57	2,58	2,49	2,80	2,72	2,56	2,46	2,39	2,40	2,26	2,52	

¹ indica que a primeira linha está vazia por falta de informações necessárias para a geração de grupos de ligação com 50 indivíduos e concentração de cinco marcas por grupo de ligação.

4.3.4. Média das distâncias entre marcas adjacentes

A média das distâncias entre marcas adjacentes ao longo de cada grupo de ligação foi obtida fazendo-se duas médias aritméticas sucessivas. A primeira média foi feita apenas entre os valores de distância encontrados dentro de cada grupo de ligação, enquanto a segunda foi obtida pela média aritmética das médias anteriormente obtidas. Nesse processo só foram utilizadas repetições que tinham 10 grupos de ligação, sendo as demais desconsideradas para efeito de análise.

Os valores de média das distâncias entre marcas adjacentes e seu desvio-padrão obtido nos vários tamanhos de população segregante utilizados no mapeamento genético estão apresentados nos Quadros 9 e 10, respectivamente.

Como as populações foram simuladas a partir de genótipos pre-determinados, a média das distâncias entre marcas adjacentes já é conhecida. Os valores a serem comparados após o mapeamento das populações segregantes deveriam ser de 20, 10, 6,67 e 5 cM nas populações geradas a partir dos genomas com níveis de saturação de 5, 10, 15 e 20 marcas, respectivamente. Como pode ser observado no Quadro 9, independentemente do tamanho da população ou do nível de saturação utilizado, todos os valores ficaram acima dos inicialmente esperados.

Em relação ao teste, foi encontrado diferença entre as médias no menor nível de saturação. Mesmo não sendo constatado diferença entre as médias dos grupos de ligação avaliados nessa saturação, as letras nas médias dos três maiores níveis de saturação representam o que também acontece em relação aos grupos de ligação nos diferentes tamanhos de população.

Considerando o teste de médias feito com $P < 0,01$, encontrou-se diferença apenas no menor nível de saturação. Nos demais níveis, independentemente do tamanho das populações, não foram constatadas diferenças significativas entre os valores das médias.

No nível de saturação de cinco marcas, a população com 100 indivíduos difere das populações, com 400, 600 e 800 indivíduos. As populações com 150 e 200 indivíduos não diferem das demais.

A falta de tendência à diminuição no tamanho da distância entre marcas adjacentes pode ser claramente vista, por exemplo, no grupo de ligação 2 na saturação com 20 marcas por grupo de ligação, em que na população com 50 indivíduos a distância entre marcas é de 5,26 cM, enquanto na população com 800 indivíduos a distância entre médias é de 5,36; como visto, ocorre pequeno aumento nos valores. É percebido que, nesse mesmo grupo de ligação, existe pequena variação nos valores, mas não mostram tendência de diminuição. Essa variação também pode ser observada nos demais grupos de ligação, em todos os níveis de saturação.

Quadro 9 – Médias das distâncias entre marcas adjacentes nas repetições que recuperaram 10 grupos de ligação, em populações F_2 , avaliados em quatro níveis de saturação

Saturação	T.P.	Grupos de Ligação										Média Geral	
		1	2	3	4	5	6	7	8	9	10		
5 marcas	50	-	-	-	-	-	-	-	-	-	-	-	¹
	100	24,62	24,20	24,34	24,72	25,64	24,23	24,23	24,53	24,97	24,58	24,61 (b)	²
	150	24,76	24,97	25,29	24,86	24,82	25,11	25,09	24,96	24,90	25,11	24,99 (ab)	
	200	24,75	25,01	25,13	24,91	25,08	24,66	25,17	24,82	25,19	25,14	24,98 (ab)	
	400	25,26	25,13	25,18	25,16	25,35	25,35	25,42	25,22	25,32	25,26	25,27 (a)	
	600	25,22	25,27	25,23	25,39	25,30	25,13	25,35	25,25	25,24	25,33	25,27 (a)	
	800	25,19	25,28	25,35	25,29	25,32	25,30	25,32	25,32	25,17	25,22	25,28 (a)	
10 marcas	50	11,44	11,55	11,39	11,14	11,34	11,70	11,45	11,37	11,45	11,50	11,43 (a)	
	100	11,50	11,53	11,46	11,40	11,39	11,36	11,50	11,35	11,38	11,55	11,44 (a)	
	150	11,45	11,50	11,50	11,35	11,40	11,34	11,34	11,46	11,38	11,36	11,41 (a)	
	200	11,38	11,49	11,32	11,42	11,48	11,43	11,42	11,39	11,49	11,44	11,43 (a)	
	400	11,41	11,37	11,41	11,41	11,39	11,43	11,43	11,43	11,44	11,42	11,41 (a)	
	600	11,40	11,38	11,42	11,42	11,42	11,44	11,43	11,41	11,41	11,38	11,41 (a)	
	800	11,44	11,37	11,44	11,41	11,44	11,41	11,40	11,40	11,44	11,45	11,42 (a)	
15 marcas	50	7,43	7,14	7,25	7,27	7,22	7,45	7,40	7,30	7,34	7,41	7,32 (a)	
	100	7,35	7,38	7,27	7,43	7,20	7,37	7,30	7,29	7,35	7,36	7,33 (a)	
	150	7,30	7,26	7,33	7,29	7,29	7,29	7,35	7,34	7,28	7,26	7,30 (a)	
	200	7,25	7,33	7,34	7,31	7,35	7,29	7,32	7,29	7,32	7,34	7,31 (a)	
	400	7,35	7,31	7,29	7,32	7,35	7,32	7,31	7,31	7,36	7,32	7,33 (a)	
	600	7,35	7,31	7,32	7,31	7,30	7,30	7,29	7,32	7,29	7,31	7,31 (a)	
	800	7,32	7,31	7,26	7,32	7,30	7,29	7,32	7,33	7,29	7,29	7,30 (a)	
20 marcas	50	5,38	5,26	5,50	5,36	5,40	5,36	5,45	5,42	5,44	5,47	5,40 (a)	
	100	5,29	5,33	5,34	5,30	5,37	5,42	5,33	5,38	5,28	5,40	5,34 (a)	
	150	5,33	5,33	5,38	5,39	5,38	5,34	5,32	5,32	5,35	5,40	5,35 (a)	
	200	5,36	5,35	5,33	5,34	5,35	5,31	5,35	5,36	5,39	5,37	5,35 (a)	
	400	5,37	5,34	5,38	5,33	5,39	5,38	5,34	5,35	5,33	5,37	5,36 (a)	
	600	5,36	5,33	5,35	5,38	5,36	5,34	5,35	5,37	5,36	5,38	5,36 (a)	
	800	5,34	5,36	5,37	5,34	5,34	5,36	5,37	5,36	5,37	5,34	5,36 (a)	

¹ indica que a primeira linha está vazia por falta de informações necessárias para a geração de grupos de ligação com 50 indivíduos e concentração de cinco marcas por grupo de ligação; T.P. = tamanho da população; ² indica que dentro dos parêntesis estão as letras correspondentes ao resultado da análise de médias pelo teste de Tukey no nível de significância de 1%.

Embora a diminuição nos valores das médias entre marcas adjacentes ocorra, nesse caso apenas em função do aumento no nível de saturação, o desvio-padrão apresenta comportamento diferente.

No Quadro 10 estão os desvios das médias das distâncias entre marcas adjacentes nas repetições que recuperaram 10 grupos de ligação, avaliados em quatro níveis de saturação. É interessante observar que, apesar dos valores muito

próximos, pode ser observado que, com o aumento do tamanho da população, parece existir tendência à diminuição na amplitude de variação nos valores das médias das distâncias entre marcas adjacentes. No nível de saturação com cinco marcas por grupo de ligação, essa tendência pode ser observada nas populações iguais e superiores a 400 indivíduos. Nas populações com 100, 150 e 200 indivíduos, essa tendência existe apenas nos grupos de ligação 1, 4, 6 e 10.

Quadro 10 – Desvio-padrão das médias das distâncias entre marcas adjacentes nas repetições que recuperaram 10 grupos de ligação, em populações F_2 , avaliados em quatro níveis de saturação

Saturação	T.P.	Grupos de Ligação										Média dos Desvios	
		1	2	3	4	5	6	7	8	9	10		
5 marcas	50	-	-	-	-	-	-	-	-	-	-	-	¹
	100	1,75	1,28	1,70	1,61	1,11	1,80	1,28	1,44	1,17	1,79	1,49	
	150	1,58	1,34	1,16	1,54	1,40	1,19	1,39	1,64	1,25	1,38	1,38	
	200	1,32	1,19	1,34	1,06	1,44	1,16	0,91	1,35	1,15	1,13	1,20	
	400	0,80	0,96	0,95	0,84	0,75	0,92	0,88	1,01	0,84	0,85	0,88	
	600	0,73	0,74	0,83	0,82	0,69	0,76	0,74	0,83	0,73	0,73	0,76	
	800	0,62	0,68	0,71	0,65	0,67	0,64	0,64	0,55	0,69	0,61	0,64	
10 marcas	50	1,11	1,03	1,10	1,22	1,32	1,06	1,08	1,26	1,12	1,18	1,14	
	100	0,73	0,92	0,82	0,79	0,74	0,83	0,80	0,87	0,78	0,73	0,80	
	150	0,70	0,67	0,65	0,69	0,57	0,69	0,64	0,68	0,61	0,66	0,65	
	200	0,51	0,58	0,51	0,53	0,56	0,52	0,57	0,55	0,56	0,56	0,54	
	400	0,40	0,36	0,38	0,33	0,36	0,38	0,36	0,40	0,37	0,34	0,36	
	600	0,29	0,33	0,31	0,31	0,27	0,32	0,32	0,32	0,27	0,29	0,30	
	800	0,28	0,32	0,26	0,29	0,29	0,30	0,30	0,28	0,25	0,29	0,28	
15 marcas	50	0,80	0,70	0,76	0,64	0,74	0,70	0,85	0,75	0,80	0,69	0,74	
	100	0,56	0,56	0,51	0,50	0,54	0,50	0,57	0,50	0,51	0,49	0,52	
	150	0,51	0,43	0,45	0,37	0,48	0,45	0,45	0,40	0,45	0,42	0,44	
	200	0,31	0,35	0,37	0,42	0,34	0,32	0,34	0,37	0,36	0,34	0,35	
	400	0,26	0,26	0,24	0,24	0,28	0,25	0,26	0,28	0,26	0,26	0,25	
	600	0,19	0,22	0,23	0,21	0,19	0,20	0,21	0,22	0,19	0,22	0,20	
	800	0,17	0,17	0,18	0,18	0,17	0,17	0,18	0,18	0,18	0,20	0,17	
20 marcas	50	0,55	0,52	0,63	0,57	0,53	0,56	0,49	0,56	0,57	0,66	0,56	
	100	0,36	0,33	0,35	0,38	0,39	0,39	0,40	0,36	0,39	0,38	0,37	
	150	0,25	0,32	0,35	0,36	0,31	0,33	0,31	0,29	0,29	0,32	0,31	
	200	0,23	0,27	0,26	0,23	0,25	0,26	0,26	0,29	0,27	0,24	0,25	
	400	0,22	0,18	0,18	0,17	0,19	0,18	0,17	0,20	0,19	0,20	0,18	
	600	0,17	0,16	0,15	0,13	0,15	0,16	0,17	0,15	0,16	0,16	0,15	
	800	0,13	0,13	0,13	0,14	0,14	0,13	0,12	0,12	0,12	0,11	0,12	

T.P. = tamanho da população; e ¹ indica que a primeira linha está vazia por falta de informações necessárias para a geração de grupos de ligação com 50 indivíduos e concentração de cinco marcas por grupo de ligação.

Nos demais níveis de saturação, apenas um dos grupos de ligação apresentou valores maiores em populações com maior número de indivíduos. Na saturação com 10 marcas, no grupo de ligação 5 as populações com 600 e 800 indivíduos apresentaram valores de desvio-padrão de 0,27 e 0,29, respectivamente. No nível de saturação com 15 marcas, no grupo de ligação 4 as populações com 150 e 200 indivíduos apresentaram desvio-padrão de 0,37 e 0,42, respectivamente. Finalmente, no maior nível de saturação e também no grupo de ligação 4, os desvios foram de 0,13 e 0,14 nas populações com 600 e 800 indivíduos, respectivamente.

Por não apresentarem nenhuma diferença significativa, as médias das distâncias entre as marcas deveriam exibir valores de desvio-padrão com comportamento semelhante, mas pelo visto é demonstrado que, apesar de o tamanho da população não influir no tamanho das médias nos três maiores níveis de saturação, existe influência na diminuição na variação da amplitude do tamanho das médias com o aumento no tamanho das populações avaliadas dentro de cada nível de saturação.

Como já comentado, segundo Hospital et al. (1992), marcadores são mais úteis quando sua posição no mapa é conhecida. Com isso, tem-se que, quanto menor a variância e, conseqüentemente, o desvio das distâncias entre marcas, melhor será a capacidade de localização e, por conseqüência, mais eficiente o trabalho de mapeamento.

Um método de mapeamento proposto por Lander e Botstein (1989) emprega pares de marcadores vizinhos para obter o máximo de informação da ligação de QTLs, dentro do segmento de cromossomo analisado. Esses autores investigaram a acurácia do método por simulação, e os resultados indicaram a existência de uma probabilidade razoável de detecção de QTL que explique pelo menos 5% da variância. Na média, um QTL com capacidade para explicar 5 ou 10% da variância é mapeado com um intervalo de 40 ou 20 cM, respectivamente. É claro que QTLs com maiores efeitos genotípicos serão localizados mais precisamente. Contudo, deve ser notado que o comprimento do intervalo é variável.

4.3.5. Variância das distâncias entre marcas adjacentes

A partir das distâncias entre as marcas adjacentes obtidas nos grupos de ligação foi estimada a variância amostral, como apresentado no Quadro 11. Como em quadros anteriores, para cada de grupo de ligação são dadas médias aritméticas, nesse caso, das variâncias obtidas em cada repetição, em que houve a formação de 10 grupos de ligação no mapeamento das populações F_2 simuladas.

Os valores de variância encontrados na análise dos mapas obtidos nas populações segregantes simulados são referentes aos erros para qualquer tamanho de população e nível de saturação dos genomas avaliados, pois os genomas utilizados na geração das populações segregantes tinham seus marcadores distribuídos de forma equidistante dentro dos 10 grupos de ligação em todos os níveis de saturação, sendo que para o genoma com nível de saturação de cinco marcas por grupo de ligação os marcadores encontravam-se distribuídos em distâncias equidistantes de 20 cM ao longo dos grupos de ligação. Para o genoma com nível de saturação de 10 marcas por grupo de ligação, os marcadores encontravam-se distribuídos à distância de 10 cM ao longo dos grupos de ligação. Para o genoma com nível de saturação de 15 marcas por grupo de ligação, os marcadores estavam distribuídos à distância de certa de 6,67 cM e, finalmente, para o genoma com nível de saturação de 20 marcas por grupo de ligação, os marcadores encontravam-se distribuídos à distância de 5 cM ao longo dos grupos de ligação.

Os valores de variância podem ser interpretados de forma que, quanto menores os valores de variância, mais equidistantes estarão distribuídas as marcas dentro dos grupos de ligação; conseqüentemente, menor é o erro. Portanto, quanto menores os valores de variância, mais próximos estarão os valores do esperado em cada nível de saturação, indicando uma boa recuperação do genoma com o mapeamento das populações segregantes.

Quadro 11 – Variância das distâncias entre médias adjacentes nas repetições que recuperaram 10 grupos de ligação, em populações F_2 , avaliados em quatro níveis de saturação

Saturação	T.P.	Grupos de Ligação										Média Geral		
		1	2	3	4	5	6	7	8	9	10			
5 marcas	50	-	-	-	-	-	-	-	-	-	-	-	-	-
	100	9,05 (a)	10,49 (a)	9,41 (a)	9,28 (a)	9,01 (a)	7,96 (a)	7,34 (ab)	10,50 (a)	8,94 (a)	9,46 (a)	9,14 (a)	2	
	150	6,48 (ab)	6,70 (ab)	7,06 (a)	7,02 (ab)	6,00 (ab)	5,22 (ab)	7,36 (a)	7,72 (ab)	7,04 (a)	5,13 (ab)	6,57 (b)		
	200	5,34 (abc)	5,58 (abc)	4,91 (ab)	3,95 (abc)	5,08 (abc)	6,74 (a)	6,17 (ab)	5,71 (ab)	4,98 (ab)	4,51 (ab)	5,30 (b)		
	400	3,28 (bcd)	3,13 (bcd)	3,14 (bc)	3,35 (bc)	3,48 (abc)	3,04 (ab)	3,38 (ab)	3,71 (bc)	3,23 (bc)	3,26 (b)	3,30 (c)		
	600	2,21 (cd)	2,10 (cd)	2,10 (bc)	2,14 (c)	2,11 (bc)	2,67 (ab)	2,19 (b)	1,88 (c)	2,53 (bc)	2,29 (b)	2,22 (d)		
10 marcas	800	1,69 (d)	1,69 (d)	1,49 (c)	1,56 (c)	1,53 (c)	1,85 (b)	1,84 (b)	1,77 (c)	1,64 (c)	1,82 (b)	1,69 (d)		
	50	11,08 (a)	10,99 (a)	11,76 (a)	10,57 (a)	10,50 (a)	10,99 (a)	11,59 (a)	11,85 (a)	11,89 (a)	11,51 (a)	11,27 (a)		
	100	6,15 (b)	5,68 (b)	5,34 (b)	5,86 (b)	5,54 (b)	6,42 (b)	5,74 (b)	5,84 (b)	5,89 (b)	6,44 (b)	5,89 (b)		
	150	3,88 (c)	3,78 (bc)	3,93 (bc)	4,28 (bc)	3,74 (bc)	4,07 (c)	4,19 (bc)	4,02 (bc)	3,59 (c)	3,81 (c)	3,93 (c)		
	200	2,91 (cd)	2,87 (cd)	2,80 (cd)	2,99 (cd)	2,86 (cd)	3,12 (cd)	2,96 (cd)	2,87 (cd)	2,96 (cd)	2,79 (cd)	2,91 (d)		
	400	1,35 (de)	1,47 (de)	1,45 (d)	1,42 (de)	1,49 (de)	1,66 (de)	1,41 (de)	1,48 (cd)	1,44 (de)	1,57 (de)	1,47 (e)		
15 marcas	600	1,04 (de)	0,92 (e)	0,98 (d)	0,91 (e)	0,98 (e)	1,02 (e)	0,96 (e)	0,98 (d)	1,01 (de)	0,95 (de)	0,97 (ef)		
	800	0,69 (e)	0,68 (e)	0,72 (d)	0,70 (e)	0,69 (e)	0,69 (e)	0,72 (e)	0,77 (d)	0,79 (e)	0,73 (e)	0,72 (f)		
	50	8,17 (a)	7,28 (a)	7,90 (a)	7,01 (a)	7,59 (a)	6,99 (a)	7,04 (a)	7,85 (a)	7,56 (a)	7,25 (a)	7,46 (a)		
	100	3,54 (b)	3,59 (b)	3,45 (b)	3,46 (b)	4,04 (b)	3,66 (b)	3,81 (b)	3,87 (b)	4,07 (b)	3,93 (b)	3,74 (b)		
	150	2,49 (bc)	2,44 (bc)	2,49 (bc)	2,24 (c)	2,41 (c)	2,52 (c)	2,56 (c)	2,22 (bc)	2,41 (c)	2,34 (c)	2,41 (c)		
	200	1,82 (cd)	1,84 (cd)	2,01 (c)	1,78 (cd)	1,76 (cd)	1,85 (cd)	1,95 (c)	1,81 (cd)	1,82 (cd)	1,87 (cd)	1,85 (d)		
20 marcas	400	0,95 (de)	0,87 (de)	0,92 (d)	0,91 (de)	0,89 (d)	0,92 (de)	0,93 (d)	0,92 (cd)	0,87 (de)	0,92 (de)	0,91 (e)		
	600	0,61 (de)	0,61 (e)	0,64 (d)	0,61 (e)	0,60 (d)	0,62 (e)	0,62 (d)	0,64 (cd)	0,62 (de)	0,61 (de)	0,62 (ef)		
	800	0,46 (e)	0,43 (e)	0,46 (d)	0,48 (e)	0,48 (d)	0,43 (e)	0,48 (d)	0,45 (d)	0,44 (e)	0,45 (e)	0,46 (f)		
	50	5,38 (a)	5,39 (a)	7,41 (a)	5,60 (a)	5,75 (a)	5,84 (a)	5,88 (a)	5,60 (a)	6,51 (a)	6,49 (a)	5,99 (a)		
	100	2,64 (b)	2,68 (b)	2,80 (b)	2,69 (b)	2,66 (b)	2,67 (b)	2,69 (b)	2,78 (b)	2,74 (b)	2,93 (b)	2,73 (b)		
	150	1,76 (c)	1,73 (c)	1,85 (b)	1,78 (bc)	1,71 (c)	1,78 (bc)	1,77 (bc)	1,81 (c)	1,81 (bc)	1,80 (bc)	1,78 (c)		
20 marcas	200	1,32 (cd)	1,41 (cd)	1,28 (b)	1,38 (cd)	1,31 (cd)	1,29 (cd)	1,36 (cd)	1,32 (cd)	1,41 (bc)	1,29 (bc)	1,34 (c)		
	400	0,68 (de)	0,71 (de)	0,66 (b)	0,67 (cd)	0,70 (de)	0,64 (cd)	0,66 (cd)	0,64 (de)	0,69 (c)	0,65 (c)	0,67 (d)		
	600	0,46 (e)	0,44 (e)	0,42 (b)	0,45 (d)	0,44 (e)	0,46 (d)	0,44 (d)	0,45 (de)	0,46 (e)	0,47 (c)	0,45 (d)		
	800	0,34 (e)	0,31 (e)	0,32 (b)	0,33 (d)	0,31 (e)	0,34 (d)	0,35 (d)	0,33 (e)	0,32 (c)	0,34 (c)	0,33 (d)		

T.P. = tamanho da população; ¹ indica que a primeira linha está vazia por falta de informações necessárias para a geração de grupos de ligação com 50 indivíduos e concentração de cinco marcas por grupo de ligação; ² indica que dentro dos parêntesis estão letras correspondentes ao resultado da análise de médias, pelo teste de Tukey no nível de significância de 1%. Nas colunas, médias seguidas de uma mesma letra não diferem entre si, pelo teste de Tukey.

Analisando o efeito do tamanho da população na menor saturação, verificou-se redução da média das variâncias em relação às distâncias entre marcas adjacentes, a não ser nos grupos de ligação 6 e 7. À medida que aumenta o tamanho da população segregante, por exemplo, no grupo 1 tem-se que o valor da variância foi de 9,05 e 1,69, nos tamanhos de população de 100 e 800 indivíduos, respectivamente (Quadro 11).

Além da redução na média das variâncias, observa-se redução na amplitude de variação dos valores de variância entre repetições, à medida que aumentou o tamanho da população segregante. Essa redução pode ser demonstrada pela redução no desvio-padrão. Como exemplo, podem-se observar os desvios nas mesmas populações citadas, em que o desvio-padrão foi de 7,06 e 1,42, nos tamanhos de população de 100 e 800 indivíduos, respectivamente (Quadro 12).

Na análise da média geral houve diferença significativa entre média dos diferentes tamanhos de população, sendo que não diferiram entre si as médias das populações de tamanho 150 e 200 e aquelas de tamanho 600 e 800.

Nos sete tamanhos de população segregantes analisados dentro do nível de saturação de 10 marcas por grupo de ligação, assim como no nível de saturação anterior, houve redução no desvio-padrão das médias das distâncias entre marcas adjacentes, à medida que aumentou o tamanho da população segregante, só que nesse caso o fato se deu em todos os grupos de ligação. Como exemplo, mantendo o grupo de ligação 1 como referência, as variâncias foram 11,08 e 0,69 nas populações com 50 e 800 indivíduos, respectivamente (Quadro 11); em relação ao desvio-padrão, os valores foram 4,81 e 0,36 nessas mesmas populações (Quadro 12). Também foi constatado redução na amplitude de variação dos valores de variância entre repetições, com o aumento no tamanho da população segregante. Essa diminuição também é constatada pela observação no comportamento do desvio-padrão (Quadro 12).

Em relação à confiabilidade, quanto menores os valores de variância dos valores entre marcas adjacentes e de desvio-padrão entre repetições, maior a confiabilidade dos dados.

Quadro 12 – Desvio-padrão das médias das variâncias das distâncias entre marcas adjacentes nas repetições que recuperaram 10 grupos de ligação, em populações F_2 , avaliados em quatro níveis de saturação

Saturação	T.P.	Grupos de Ligação										
		1	2	3	4	5	6	7	8	9	10	
5 marcas	50	-	-	-	-	-	-	-	-	-	-	¹
	100	7,06	7,56	5,93	7,62	8,87	6,39	6,48	8,29	5,28	8,64	
	150	5,47	5,06	4,19	4,58	4,81	5,38	7,50	5,84	5,35	4,49	
	200	3,93	3,86	3,28	3,51	4,16	4,76	4,00	4,07	3,39	3,27	
	400	2,29	2,65	2,61	2,88	2,91	2,20	2,36	2,70	2,29	2,15	
	600	1,67	1,71	1,57	1,82	1,76	2,13	1,67	1,47	1,92	2,01	
	800	1,42	1,36	1,45	1,23	1,11	1,47	1,30	1,27	1,38	1,43	
10 marcas	50	4,81	5,69	6,54	5,61	4,94	5,45	5,19	7,52	5,96	5,57	
	100	3,22	2,61	2,80	2,81	3,04	3,22	2,92	4,11	2,91	3,09	
	150	1,88	1,74	1,85	1,92	1,72	2,01	2,21	1,86	1,84	1,93	
	200	1,48	1,29	1,18	1,32	1,55	1,50	1,36	1,44	1,50	1,51	
	400	0,65	0,62	0,74	0,72	0,73	0,91	0,70	0,69	0,84	0,78	
	600	0,52	0,51	0,48	0,45	0,44	0,51	0,46	0,44	0,45	0,48	
	800	0,36	0,41	0,36	0,34	0,38	0,32	0,31	0,38	0,44	0,34	
15 marcas	50	4,03	3,36	2,87	2,90	3,53	3,05	2,58	4,73	3,11	3,07	
	100	1,29	1,32	1,44	1,30	2,94	1,53	1,49	2,75	2,63	2,70	
	150	0,87	0,95	0,85	0,78	0,91	0,92	0,98	0,89	0,98	0,93	
	200	0,61	0,83	0,81	0,65	0,60	0,90	0,78	0,72	0,64	0,76	
	400	0,41	0,36	0,35	0,41	0,35	0,39	0,35	0,38	0,35	0,38	
	600	0,26	0,22	0,22	0,26	0,22	0,24	0,20	0,21	0,25	0,20	
	800	0,19	0,15	0,19	0,18	0,20	0,16	0,20	0,18	0,19	0,17	
20 marcas	50	1,80	2,26	9,42	3,94	2,01	3,61	3,51	2,86	5,89	6,23	
	100	0,95	0,89	0,98	0,87	0,91	0,89	0,80	0,94	0,88	0,98	
	150	0,63	0,50	0,65	0,60	0,57	0,72	0,50	0,61	0,55	0,65	
	200	0,45	0,50	0,38	0,49	0,53	0,41	0,51	0,37	0,48	0,39	
	400	0,24	0,24	0,23	0,23	0,22	0,22	0,23	0,23	0,24	0,21	
	600	0,15	0,15	0,12	0,16	0,12	0,15	0,12	0,14	0,15	0,15	
	800	0,12	0,11	0,10	0,10	0,10	0,11	0,12	0,11	0,10	0,11	

¹ indica que a primeira linha está vazia por falta de informações necessárias para a geração de grupos de ligação com 50 indivíduos e concentração de cinco marcas por grupo de ligação.

A redução nos valores de variância com o aumento do tamanho da população é evidenciada pelas diferenças significativas dadas pelo teste de médias, em que as menores médias estão associadas às maiores populações em todos os grupos de ligação dentro do nível de saturação de 10 marcas por grupo de ligação. As médias gerais apresentam diferença significativa nos tamanhos de população, evidenciando-se que as menores populações apresentam maiores

valores de variância das distâncias entre marcas adjacentes, portanto levando à menor precisão no mapeamento genético.

Em relação ao nível de saturação de 15 marcas por grupo de ligação, os valores de variância média das distâncias entre marcas adjacentes também diminuíram com o aumento no tamanho da população, como aconteceu com as populações nos níveis de saturação com 5 e 10 marcas por grupo de ligação. Como exemplo, nesse nível de saturação pode ser observada a diminuição nos valores de variância do grupo de ligação 1, passando de 8,17 na população com 50 indivíduos para 0,46 na população com 800 indivíduos (Quadro 11); em relação ao desvio-padrão, têm-se 4,03 e 0,19 nas populações com 50 e 800 indivíduos, respectivamente (Quadro 12).

Assim como nos níveis de saturação comentados anteriormente, no nível de saturação com 20 marcas por grupo de ligação os valores das variâncias diminuíram com o aumento no tamanho das populações, como indicado pelo teste de médias (Quadro 11). Os valores de desvio-padrão (Quadro 12) também indicaram diminuição com o aumento no número de indivíduos na população.

Em resumo, o que pôde ser observado em relação às médias das variâncias foi que, à medida que o tamanho da população aumentava, o valor da variância diminuía. Além disso, os valores de variância são afetados pelo nível de saturação utilizado, em que se nota que, reduzindo o nível de saturação, os valores das médias das variâncias aumentam. Em um mesmo tamanho de população e aumentando apenas o nível de saturação, pode-se observar que os valores de variância diminuem, variando de 9,05, na população com 100 indivíduos no nível de saturação com cinco marcas por grupo de ligação, até 2,64 no nível de saturação com 20 marcas por grupo de ligação.

4.3.6. Estresse

O estresse foi utilizado para expressar o grau de concordância dos valores de distância entre cada par de marcas adjacentes nos grupos de ligação simulados em relação às distâncias nos respectivos pares de marcas no genoma de referência.

Deve-se ter atenção ao analisar valores de estresse por estes apresentarem significados diferentes em níveis de saturação diferentes, ou seja, devem ser comparados níveis de estresse entre genomas com o mesmo nível de saturação, uma vez que valores iguais de porcentagem em níveis de saturação diferentes têm significados diferentes.

Para obtenção do estresse médio foi feita uma média aritmética dos valores de estresse obtidos nas repetições em que houve a formação de 10 grupos de ligação no mapeamento genético. Os valores obtidos são apresentados no Quadro 13, juntamente com o resultado de comparações feitas com teste de médias.

Analisando os efeitos do tamanho da população dentro do nível de saturação de genoma com cinco marcas por grupo de ligação, apesar do menor número de marcas nota-se que em alguns dos grupos de ligação já é observável a tendência à redução dos valores de estresse médio à medida que aumenta o tamanho da população segregante. Essa tendência pode ser mais bem observada nos valores das médias gerais de estresse representados na última coluna do Quadro 13, em que valores de estresse médio passam de 11,05 para 5,02%, considerando-se as populações com 100 e 800 indivíduos, respectivamente.

De acordo com os valores de estresse apresentados, ainda não é claro o efeito da redução da amplitude da variação dos valores de estresse com o aumento do tamanho da população segregante, nos demais níveis de saturação. Provavelmente isso ocorra em razão do pouco número de marcadores e repetições avaliadas. Em relação a isso, nos demais níveis de saturação as populações apresentam comportamento mais bem definido, como pode ser visto no Quadro 14.

Quadro 13 – Valores de estresse médio (%) em razão do tamanho da população (T.P.) e do nível de saturação do genoma em populações F₂

Saturação	T.P.	Grupos de Ligação										Média (%)												
		1	2	3	4	5	6	7	8	9	10													
5 marcas	50	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	100	10,52 (a)	14,88 (a)	8,54 (ab)	13,87 (a)	11,32 (a)	12,19 (a)	8,68 (a)	10,87 (ab)	9,34 (ab)	10,25 (a)	11,05 (a)	2											
	150	10,05 (a)	9,69 (a)	10,72 (a)	10,58 (a)	9,87 (a)	8,87 (a)	9,28 (a)	12,04 (a)	9,51 (a)	9,00 (a)	9,96 (a)												
	200	8,39 (a)	9,31 (a)	8,49 (ab)	7,60 (ab)	8,71 (a)	9,18 (a)	8,64 (a)	8,89 (ab)	8,63 (b)	7,93 (a)	8,58 (a)												
	400	6,73 (a)	6,86 (ab)	6,59 (b)	6,81 (ab)	6,87 (a)	6,83 (a)	7,05 (a)	7,47 (b)	6,74 (b)	6,76 (a)	6,87 (b)												
	600	5,62 (a)	5,58 (b)	5,73 (b)	5,80 (ab)	5,46 (a)	6,04 (a)	6,64 (a)	5,46 (b)	5,94 (b)	5,68 (a)	5,80 (c)												
800	4,93 (a)	5,08 (b)	4,88 (b)	4,92 (b)	4,97 (a)	5,15 (a)	5,23 (a)	4,98 (b)	4,95 (b)	5,08 (a)	5,02 (c)													
10 marcas	50	29,44 (a)	28,85 (a)	29,77 (a)	28,44 (a)	29,01 (a)	29,49 (a)	29,61 (a)	29,48 (a)	29,93 (a)	30,19 (a)	29,42 (a)												
	100	21,78 (b)	21,66 (b)	20,62 (b)	21,35 (b)	20,56 (b)	22,23 (b)	21,23 (b)	20,87 (b)	21,36 (b)	22,25 (b)	21,39 (b)												
	150	17,69 (c)	17,43 (c)	17,75 (bc)	18,26 (bc)	17,01 (c)	17,83 (c)	17,88 (c)	17,92 (bc)	16,72 (c)	17,17 (c)	17,57 (c)												
	200	15,01 (c)	15,30 (c)	14,77 (c)	15,37 (c)	15,12 (c)	15,59 (c)	15,33 (c)	14,99 (c)	15,43 (c)	14,88 (c)	15,18 (d)												
	400	10,64 (d)	10,87 (d)	10,83 (d)	10,64 (d)	10,85 (d)	11,47 (d)	10,74 (d)	11,02 (d)	10,78 (d)	11,16 (d)	10,90 (e)												
	600	9,20 (d)	8,73 (d)	9,07 (d)	8,82 (d)	9,08 (d)	9,30 (de)	9,08 (d)	9,14 (d)	9,10 (d)	8,87 (d)	9,04 (f)												
800	7,91 (d)	7,73 (d)	8,03 (d)	7,94 (d)	7,91 (d)	7,91 (e)	8,01 (d)	8,15 (d)	8,25 (d)	8,17 (d)	8,00 (f)													
15 marcas	50	39,51 (a)	36,32 (a)	38,80 (a)	35,98 (a)	37,58 (a)	36,60 (a)	37,17 (a)	37,45 (a)	38,29 (a)	37,15 (a)	37,48 (a)												
	100	26,77 (b)	26,94 (b)	25,94 (b)	26,23 (b)	27,11 (b)	26,71 (b)	27,31 (b)	26,37 (b)	27,29 (b)	26,90 (b)	26,76 (b)												
	150	22,37 (c)	21,81 (c)	22,24 (c)	20,84 (c)	21,97 (c)	22,25 (c)	22,51 (c)	20,91 (c)	21,72 (c)	21,32 (c)	21,79 (c)												
	200	18,68 (c)	18,89 (c)	19,87 (c)	18,95 (c)	18,72 (c)	18,75 (c)	19,46 (c)	18,86 (c)	19,00 (c)	19,09 (c)	19,03 (d)												
	400	13,94 (d)	13,27 (d)	13,54 (d)	13,53 (d)	13,67 (d)	13,56 (d)	13,76 (d)	13,69 (d)	13,48 (d)	13,66 (d)	13,61 (e)												
	600	11,36 (de)	11,35 (d)	11,61 (de)	11,26 (d)	11,05 (de)	11,26 (de)	11,33 (de)	11,65 (de)	11,20 (de)	11,33 (de)	11,34 (f)												
800	9,83 (e)	9,64 (d)	9,67 (e)	10,11 (d)	9,96 (e)	9,60 (e)	10,10 (e)	9,92 (e)	9,58 (e)	9,73 (e)	9,81 (g)													
20 marcas	50	43,58 (a)	43,15 (a)	44,14 (a)	42,69 (a)	44,88 (a)	43,73 (a)	44,05 (a)	43,84 (a)	43,93 (a)	43,57 (a)	43,75 (a)												
	100	30,73 (b)	31,04 (b)	31,70 (b)	31,25 (b)	31,15 (b)	31,47 (b)	31,38 (b)	31,73 (b)	31,47 (b)	32,64 (b)	31,46 (b)												
	150	25,09 (c)	25,24 (c)	26,18 (c)	25,84 (c)	25,07 (c)	25,44 (c)	25,45 (c)	25,57 (c)	25,69 (c)	25,78 (c)	25,53 (c)												
	200	21,92 (c)	22,70 (c)	21,71 (d)	22,32 (c)	21,89 (c)	21,70 (c)	22,25 (c)	22,23 (c)	22,81 (c)	21,80 (d)	22,13 (d)												
	400	16,13 (d)	16,21 (d)	15,78 (e)	15,72 (d)	16,33 (d)	15,64 (d)	15,63 (d)	15,47 (d)	15,97 (d)	15,75 (e)	15,86 (e)												
	600	13,36 (de)	12,97 (de)	12,74 (ef)	13,21 (de)	13,07 (de)	13,19 (de)	13,14 (de)	13,20 (de)	13,40 (de)	13,48 (ef)	13,18 (f)												
800	11,41 (e)	11,10 (e)	11,40 (f)	11,42 (e)	11,12 (e)	11,50 (e)	11,73 (e)	11,33 (e)	11,32 (e)	11,43 (f)	11,38 (g)													

T.P. = tamanho da população; ¹ indica que a primeira linha está vazia por falta de informações necessárias para a geração de grupos de ligação com 50 indivíduos e concentração de cinco marcas por grupo de ligação; ² indica que dos parêntesis estão letras correspondentes ao resultado da análise de médias feita pelo teste de Tukey a 1%. Nas colunas, médias seguidas de uma mesma letra não diferem entre si, pelo teste de Tukey.

Quadro 14 – Desvio-padrão dos valores de estresse médio nas repetições que recuperaram 10 grupos de ligação, em populações F₂, avaliados em quatro níveis de saturação

Saturação	T.P.	Grupos de Ligação										
		1	2	3	4	5	6	7	8	9	10	
5 marcas	50	-	-	-	-	-	-	-	-	-	-	¹
	100	4,48	4,16	0,65	0,90	2,68	3,35	4,43	3,58	4,47	3,87	
	150	2,96	2,85	2,06	2,97	2,44	3,17	4,50	2,34	3,02	3,60	
	200	2,96	3,07	2,79	2,43	2,95	3,01	2,42	3,21	2,47	2,94	
	400	2,29	2,46	2,78	2,49	2,53	2,28	2,37	2,49	2,16	2,25	
	600	2,07	2,04	1,93	2,30	1,79	2,24	10,00	2,17	2,06	2,27	
	800	1,66	1,81	2,03	1,68	1,63	1,97	1,74	1,73	1,83	1,69	
10 marcas	50	6,01	7,02	8,26	7,91	7,03	6,96	6,50	7,51	7,13	7,23	
	100	5,20	4,98	5,04	4,89	5,34	5,52	5,37	5,75	5,04	5,47	
	150	4,23	4,33	4,05	4,41	3,76	4,42	4,51	4,18	4,05	4,66	
	200	3,62	3,65	3,14	3,30	3,85	3,73	3,65	3,84	3,68	3,93	
	400	2,42	2,29	2,70	2,62	2,69	3,05	2,54	2,72	3,06	2,62	
	600	2,26	2,37	2,23	2,05	1,80	2,17	2,14	1,99	2,03	2,12	
	800	1,97	2,12	1,75	1,77	2,06	1,79	1,74	1,88	2,12	1,94	
15 marcas	50	9,64	8,02	7,19	7,62	7,92	7,37	7,02	7,54	7,73	7,94	
	100	5,07	4,92	5,3	4,73	5,65	5,31	5,44	5,48	5,30	4,75	
	150	4,39	4,05	3,98	3,80	3,95	4,15	4,37	4,13	4,43	4,11	
	200	3,32	4,33	3,96	3,80	3,39	4,29	3,78	3,74	3,26	3,98	
	400	2,75	2,88	2,55	2,97	2,76	3,00	2,59	2,79	2,53	2,86	
	600	2,36	2,09	2,05	2,29	2,03	2,31	1,82	1,99	2,25	2,01	
	800	2,04	1,66	1,99	1,91	2,07	1,52	2,18	1,89	2,00	2,04	
20 marcas	50	6,90	8,05	7,70	7,49	8,23	9,23	7,75	7,92	7,96	8,88	
	100	5,67	5,09	5,48	5,13	5,46	4,96	4,94	5,54	5,02	5,59	
	150	4,48	3,86	4,71	4,62	4,45	5,32	3,92	4,36	3,89	4,66	
	200	3,92	4,05	3,48	4,01	4,02	3,56	4,02	3,33	4,05	3,38	
	400	3,00	2,60	2,88	2,71	2,67	2,75	2,96	2,96	3,06	2,57	
	600	2,01	2,36	2,03	2,24	1,88	2,26	1,99	2,14	2,27	2,28	
	800	1,91	2,08	1,80	1,84	1,88	2,01	2,02	1,95	1,81	1,87	

¹ indica que a primeira linha está vazia por falta de informações necessárias para geração de grupos de ligação com 50 indivíduos e concentração de cinco marcas por grupo de ligação.

Agora, analisando os efeitos do tamanho da população dentro do nível de saturação de genoma com 10 marcas por grupo de ligação, nota-se que em todos os grupos de ligação os valores de estresse das populações com 50 indivíduos diferem de todas as outras com maior número de indivíduos. Com 100 indivíduos, essa diferença significativa ocorre na maioria dos grupos de ligação, a não ser nos grupos 3, 4 e 8. Os demais tamanhos de população podem ser divididos em grupos com dois ou três tamanhos de população, cujos valores

médios de estresse não diferem entre si, pelo teste de médias. Aqui a tendência à redução dos valores de estresse médio, à medida que aumenta o tamanho da população segregante, é mais bem definida, o que não ocorre nas populações com menor nível de saturação. Em relação à média geral dos valores de estresse, apresentados na última coluna do Quadro 13, existe diferença significativa em relação a quase todos os tamanhos de população, a não ser nos dois últimos, com 600 e 800 indivíduos, com valores de estresse de 9,04 (f) e 8,00% (f), respectivamente.

Em relação à redução na amplitude da variação dos valores médios de estresse com o aumento no tamanho da população segregante, a não ser nos grupos de ligação 2, 5 e 9, ocorre diminuição na amplitude de variação dos valores de estresse com o aumento do tamanho das populações segregantes, como se pode ver no Quadro 14. Essa melhor definição na redução da amplitude se dá provavelmente em razão do maior número de marcadores e repetições em relação ao nível de saturação anterior.

Quanto ao nível de saturação com 15 marcas por grupo de ligação, à semelhança do que acontece no caso anterior, os valores médios de estresse das populações com 50 indivíduos, em todos os grupos de ligação, difere das médias das populações maiores. Nesse nível de saturação, os valores de estresse das populações com 100 indivíduos diferem das médias dos demais tamanhos de população analisados. Os valores médios de estresse dos demais tamanhos de população podem ser classificados em dois ou três grupos, dependendo do grupo de ligação, dentro dos quais os valores das médias dos valores médios de estresse não diferem significativamente pelo teste de Tukey. Observou-se ainda que o número de valores médios de estresse, cujas médias não diferem significativamente, aumentou. Esse aumento é mais um indicativo da diminuição nos valores médios de estresse com o incremento do tamanho das populações segregantes. Os únicos grupos de ligação que não apresentaram cinco índices relativos ao teste de médias foram os grupos 2 e 4, apresentando apenas quatro índices. Agora, em relação às médias gerais dos valores médios de estresse, última coluna do Quadro 13, todos os valores diferiram entre si pelo teste de

média, saindo de 37,48% (a) para população com 50 indivíduos para 9,81% (g) para aquela com 800; nesse caso, pode ser visto claramente que o tamanho da população influenciou a diminuição dos valores médios de estresse nesse nível de saturação do genoma.

A não ser por alguns casos apresentados em negrito no Quadro 14, os valores de desvio-padrão diminuem com o aumento do tamanho das populações analisadas; com isso, é vista a diminuição da amplitude de variação dos valores de estresse nesse nível de saturação.

Finalmente, no nível de saturação com 20 marcas por grupo de ligação o comportamento dos valores médios de estresse foi semelhante ao observado nos níveis de saturação com 10 e 15 marcas por grupo de ligação, nas populações com 50 e 100 indivíduos, diferindo entre si das populações maiores, em relação aos valores médios de estresse. Nesse nível de saturação foi observado que, nos grupos de ligação 3 e 10, os valores médios de estresse poderiam ser separados em mais um grupo de médias (f) quando se avaliam os grupos de ligação. No nível de saturação com 20 marcas por grupo de ligação, também ficou evidenciada a diminuição nos valores médios de estresse. Com o aumento no número de indivíduos nas populações, os valores médios de estresse variaram de 43,75% a 11,38% nas populações com 50 e 800 indivíduos, respectivamente, como pode ser visto na última coluna do Quadro 13, em que estão representadas as médias gerais dos valores médios de estresse.

Como comentado, a redução nos valores médios de estresse com o aumento no tamanho das populações dentro de cada nível de saturação foi evidenciada pelas diferenças significativas obtidas com a utilização do teste de comparação das médias em todos os grupos de ligação. Essa queda foi bem nítida principalmente nas médias gerais de estresse nos dois maiores níveis de saturação, em que todas as médias diferiram entre si, pelo teste de Tukey a 1%.

Pode-se concluir, pelos dados do Quadro 13, que os menores valores de estresse estavam associados às populações de maior tamanho e os maiores valores de estresse, às de menor tamanho. Como pode ser observado, também nesse quadro, além da redução nos valores de estresse médio em cada nível de

saturação, ocorreu queda na amplitude de variação dos valores de estresse entre repetições, à medida que aumentou o tamanho das populações segregantes (Quadro 14). No geral, a redução nos valores de desvio-padrão pode ser vista na maioria dos grupos de ligação nos quatro níveis de saturação, como no grupo de ligação 3, no nível de saturação de 10 marcas por grupo de ligação, em que os valores de desvio foram 8,26 e 1,75, nas populações com 50 e 800 indivíduos, respectivamente. Embora existam casos em que não houve redução na quantidade de desvios com o aumento no tamanho das populações com 600 (1,88) e 800 (1,88) indivíduos no grupo de ligação 5 no último nível de saturação ou o aumento no desvio, assim como aconteceu nas populações com 150 (3,89) e 200 (4,05) indivíduos no grupo de ligação 9, isso não invalida a conclusão anterior.

No Quadro 14 são apresentados os valores de desvio-padrão dos valores médios de estresse em quatro níveis de saturação e sete tamanhos de população.

4.4. Índice de informatividade e desvio-padrão em populações de retrocruzamento

No Quadro 15 estão apresentados os valores do índice de informatividade e desvio-padrão das estimativas das quatro frequências de recombinação (0,05; 0,10; 0,0667; e 0,20), nos sete tamanhos de população de retrocruzamento simulados neste trabalho.

Ao ser analisado o efeito do tamanho da população e não sendo considerada a estimativa da frequência de recombinação, verifica-se que o índice de informatividade será maior quanto maior for o tamanho da população avaliada. Como exemplo, podem-se observar os valores do índice de informatividade relacionado à frequência de recombinação de 0,05, que foram de 1052,63 e 16842,10 nas populações com 50 e 800 indivíduos, respectivamente.

Quadro 15 – Estimativa do índice de informatividade, variância e desvio-padrão representados de sete tamanhos de população de retrocruzamento, em quatro níveis de saturação

Frequência de Recombinação	T.P.	Índice de Informatividade	Desvio-Padrão
0,20 (5 marcas)	50	312,5	0,0565
	100	625	0,0400
	150	937,5	0,0326
	200	1250	0,0282
	400	2500	0,0200
	600	3750	0,0163
	800	5000	0,0141
0,10 (10 marcas)	50	555,55	0,0424
	100	1111,11	0,0300
	150	1666,66	0,0244
	200	2222,22	0,0212
	400	4444,44	0,0150
	600	6666,66	0,0122
	800	8888,88	0,0106
~0,0667(15 marcas)	50	803,57	0,0352
	100	1607,14	0,0249
	150	2410,71	0,0203
	200	3214,28	0,0176
	400	6428,57	0,0124
	600	9642,85	0,0101
	800	12857,14	0,0088
0,05 (20 marcas)	50	1052,63	0,0308
	100	2105,26	0,0217
	150	3157,89	0,0177
	200	4210,52	0,0154
	400	8421,05	0,0108
	600	12631,57	0,0088
	800	16842,10	0,0077

T.P. = tamanho da população.

Como a variância é inversamente relacionada ao índice de informatividade, $V(r) = \frac{1}{I(r)}$, o seu valor diminui, diminuindo também o desvio-padrão, à medida que o valor do índice de informatividade aumenta. Ainda considerando a frequência de recombinação de 0,05, é visto, no Quadro 15, que o valor do desvio-padrão varia de 0,0308 a 0,0077, nas populações com 50 e 800 indivíduos, respectivamente.

Ao analisar o efeito da magnitude da estimativa da frequência de recombinação, pode-se dizer que, quanto menor a estimativa da frequência de recombinação, maior o valor do índice de informatividade e menores os valores de desvio-padrão, considerando um mesmo tamanho de população. Essas relações são mais bem entendidas quando, por exemplo, dentro de populações com 50 indivíduos, analisadas as estimativas de frequências de recombinação de 0,05; 0,0667; 0,10; e 0,20, em que os valores do índice de informatividade são, respectivamente, 1052,63; 803,57; 555,55; e 312,5, os valores de desvios são 0,0308; 0,0352; 0,0424; e 0,0565.

Com base no que pode ser observado no Quadro 15, quanto maior o número de indivíduos genotipados para a construção de mapas genéticos, maior a confiabilidade dos resultados e, quanto maior a perspectiva de saturação do mapa por marcadores, maior a confiabilidade das estimativas de frequência de recombinação obtida e dos mapas construídos a partir dessas estimativas.

Entretanto, como as capacidades técnica e financeira de laboratórios são limitadas, é fundamental estabelecer um número mínimo ótimo para que o mapa gerado venha a ser de utilidade para o melhoramento.

4.5. Recuperação de grupos de ligação em populações de retrocruzamento

Assim como nas populações F_2 , o número de grupos de ligação esperado no processo de mapeamento das populações de retrocruzamento simuladas seria de 10 grupos de ligação, que é o número original de grupos de ligação usado como base para simulação das populações. O número de grupos de ligação obtido no mapeamento das populações de retrocruzamento variou em maior número de grupos de ligação, de acordo com o nível de saturação do genoma e com o tamanho da população utilizada para geração de cada população simulada (Figura 12).

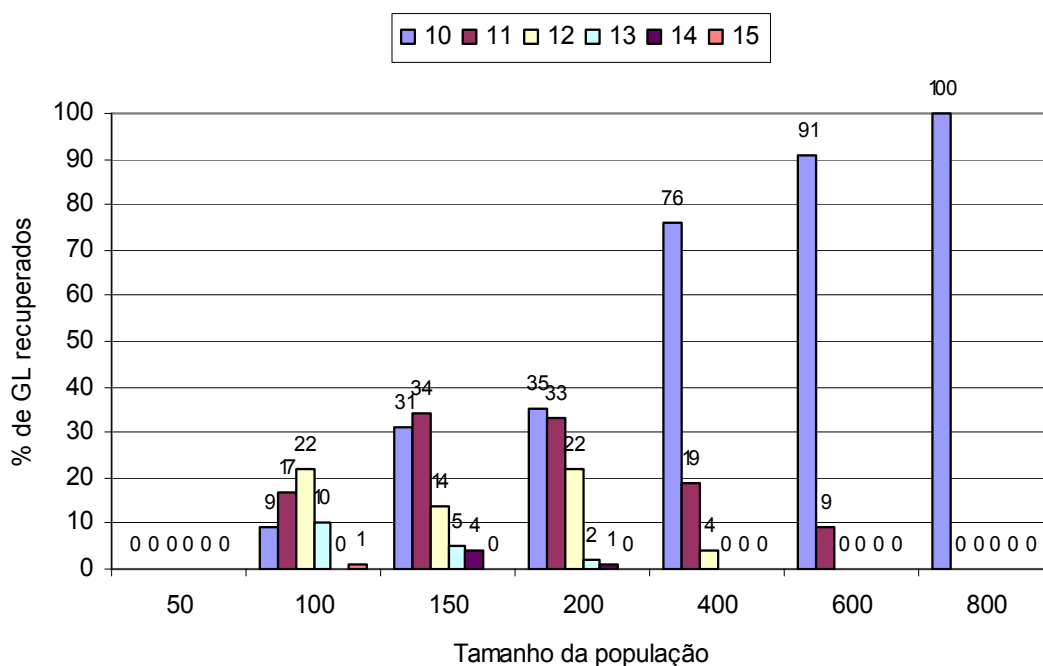


Figura 12 – Da esquerda para a direita, número de 10, 11, 12, 13, 14 e 15 grupos de ligação obtidos nas saturações de cinco marcas em populações de retrocruzamento com 50, 100, 150, 200, 400, 600 e 800 indivíduos.

4.5.1. Saturação de grupo de ligação

Na Figura 12 é apresentada a porcentagem de grupos de ligação recuperados a partir do genoma original, utilizando-se saturação de 20 cM ou cinco marcas por grupo de ligação em populações de retrocruzamento. Foram encontrados grupos de ligação variando de 10 a 15, sendo o número de repetições obtido com os diferentes números de grupos de ligação.

Com a saturação de cinco marcas por grupo de ligação em populações de retrocruzamento, também não foi possível a recuperação de nenhum grupo de ligação em nenhuma das repetições da simulação com a população de 50 indivíduos, à semelhança do que ocorreu com a população F_2 . Assim, fica evidente que esse tamanho de população de retrocruzamento é inadequado para mapeamento genético.

Ao contrário do que aconteceu com populações F_2 , foi recuperado o conjunto inicial de 10 grupos de ligação na população com 800 indivíduos em

todas as repetições; nas demais populações, nenhuma recuperou o conjunto inicial de 10 grupos de ligação em 100% das repetições simuladas, estando a população com 600 indivíduos mais próxima de atingir esse objetivo, recuperando os 10 grupos de ligação em 91 das 100 repetições.

Na população com 100 indivíduos, nove repetições recuperaram os 10 grupos de ligação, porém 17 apresentaram 11 grupos de ligação, 22 exibiram 12 grupos de ligação, 10 mostraram 13 grupos de ligação e 1 repetição apresentou 15 grupos de ligação. Nessa população houve incidência de marcas não ligadas de 11,25%, em relação ao número de marcas totais na população (Figura 12).

Na população com 150 indivíduos, 31 repetições recuperaram 10 grupos de ligação, mas 34 apresentaram 11 grupos de ligação, 14 mostraram 12 grupos de ligação, 5 exibiram 13 grupos de ligação e 4 repetições apresentaram 14 grupos de ligação. Em relação ao número de marcas não-ligadas, houve redução na população com 100 indivíduos, cuja porcentagem passou de 11,25 para 8,09%.

Na população com 200 indivíduos, 35 repetições recuperaram 10 grupos de ligação, porém 33 apresentaram 11 grupos de ligação, 22 exibiram 12 grupos de ligação, 2 mostraram 13 grupos de ligação e 1 repetição apresentou 14 grupos de ligação e uma incidência de marcas não-ligadas de 5,50%.

Na população com 400 indivíduos houve recuperação de 10 grupos de ligação em 76 repetições, mas 11 grupos de ligação em 19 repetições e 12 grupos de ligação em 4 repetições e um total de 1,49% de marcas não-ligadas.

Na população com 600 indivíduos houve a recuperação de 10 grupos de ligação em 91 repetições, mas apresentaram 11 grupos de ligação em 9 repetições e 0,54% de marcas não-ligadas.

Na população com 800 indivíduos ocorreu a recuperação de 10 grupos de ligação em todas as repetições. Além disso, foi a população com a porcentagem mais baixa de marcas não-ligadas, ou seja, apenas 0,02%.

A não ser pela população com 800 indivíduos, não foram recuperados grupos de ligação em todas as repetições, acontecendo isso em 41 repetições na população com 100 indivíduos, 12 repetições na população com 150 indivíduos, 7 repetições naquela com 200 indivíduos e 1 repetição na população com 400 indivíduos. Nas populações com 600 e 800 indivíduos houve a formação de grupos de ligação em todas as repetições. Aqui também se notou a tendência de aumento da obtenção de 10 grupos de ligação com a elevação do número de indivíduos da população.

Observa-se no Quadro 16 que, na saturação de cinco marcas por grupo de ligação, existe diminuição na porcentagem de marcas não-ligadas, que passa de 11,25 para 0,02% quando o número de indivíduos da população passa de 50 para 800 indivíduos.

Considerando a saturação de 10 marcas por grupo de ligação, na população com 50 indivíduos houve a recuperação de 10 grupos de ligação em 64 repetições, mas 24 repetições apresentaram 11 grupos de ligação (Figura 13). Na população com 100 indivíduos ocorreu a recuperação de 10 grupos de ligação em 89 repetições. Já nas populações com 150, 200, 400, 600 e 800 indivíduos houve a recuperação de 10 grupos de ligação em todas as repetições (Figura 14).

A única população a apresentar marca não-ligada foi a com 50 indivíduos, com uma porcentagem de 0,87% de marcas não-ligadas (Quadro 16).

Na saturação de 15 marcas por grupo de ligação e considerando inicialmente a população com 50 indivíduos, houve recuperação de 10 grupos de ligação em 78 repetições, porém com 11 grupos de ligação nas quatro repetições (Figura 13). Na população com 100 indivíduos houve a recuperação de 10 grupos de ligação em 95 repetições (Figura 14), enquanto nas demais populações ocorreu a recuperação de todos os grupos de ligação em todas as repetições, e apenas na população com 50 indivíduos foi constatada a existência de marcas não-ligadas (0,1056%).

Quadro 16 – Saturação e porcentagem de marcas não-ligadas em relação ao número de repetições avaliadas em sete tamanhos de população de retrocruzamento

Saturação	T.P.	MNL	Rep	%MNL
5 marcas	50	-	-	- ¹
	100	332	59	11,2542
	150	356	88	8,0909
	200	256	93	5,5053
	400	74	99	1,4949
	600	27	100	0,54
	800	1	100	0,02
10 marcas	50	76	87	0,8735
	100	0	89	0
	150	0	100	0
	200	0	100	0
	400	0	100	0
	600	0	100	0
	800	0	100	0
15 marcas	50	13	82	0,1056
	100	0	95	0
	150	0	100	0
	200	0	100	0
	400	0	100	0
	600	0	100	0
	800	0	100	0
20 marcas	50	2	86	0,0116
	100	0	93	0
	150	0	100	0
	200	0	100	0
	400	0	100	0
	600	0	100	0
	800	0	100	0

¹ indica que a primeira linha está vazia por falta de informações necessárias para a geração de grupos de ligação com 50 indivíduos e concentração de cinco marcas por grupo de ligação; saturação = número de marcas por grupo de ligação, do genoma simulado; T.P. = tamanho da população; MNL = número de marcas não-ligadas; Rep = número de repetições com 10 grupos de ligação; e %MNL = porcentagem de marcas não-ligadas em relação ao número de repetições avaliadas.

No último nível de saturação houve recuperação de 10 grupos de ligação nas 84 repetições e 11 grupos de ligação em 1 repetição, na população com 50 indivíduos, e 0,01% de marcas não-ligadas (Figura 13). Na população com 100 indivíduos ocorreu a recuperação de 10 grupos de ligação nas 93 repetições (Figura 14). Assim como nas saturações de 10 e 15 marcas por grupo de ligação, as populações com 150, 200, 400, 600 e 800 indivíduos recuperaram 10 grupos de ligação em todas as repetições.

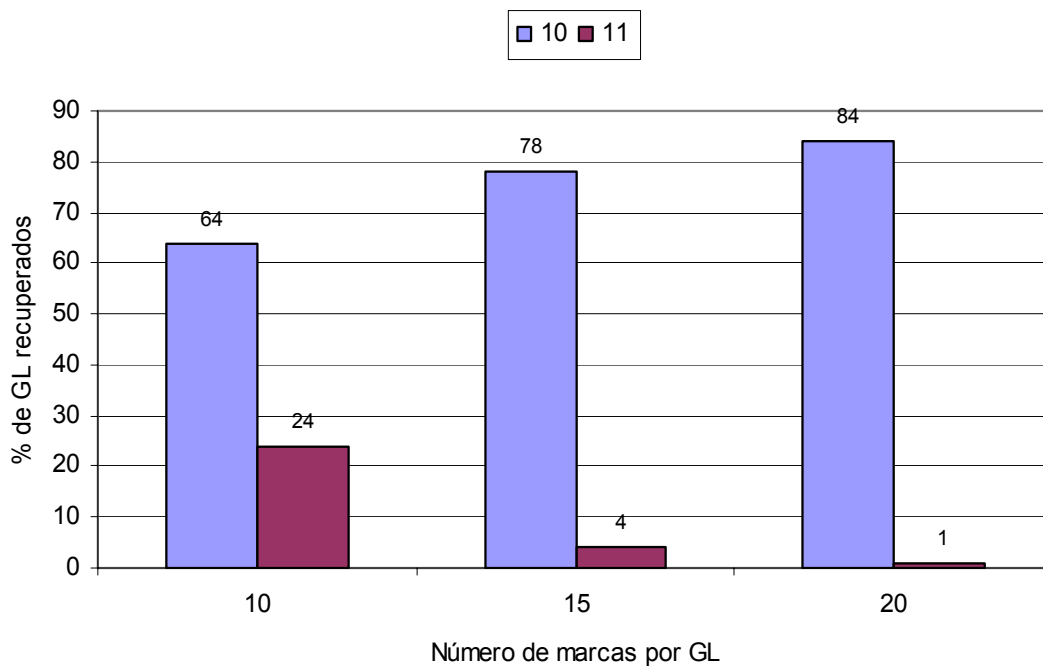


Figura 13 – Da esquerda para a direita, número de 10 e 11 grupos de ligação obtidos nas saturações de 10, 15 e 20 marcas em populações de retrocruzamento com 50 indivíduos.

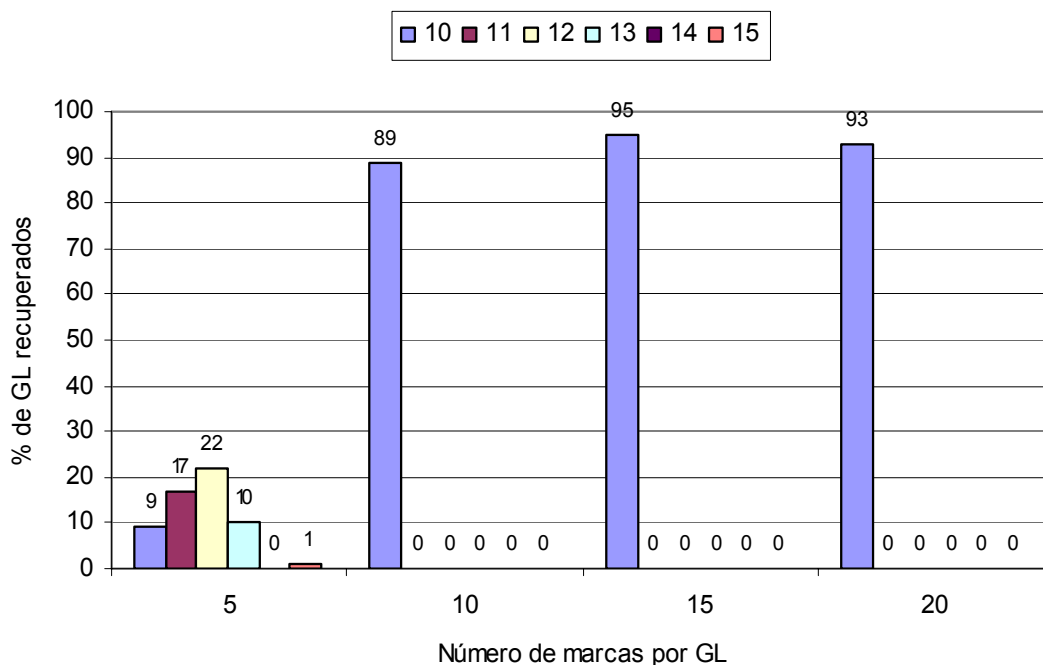


Figura 14 – Da esquerda para a direita, número de 10, 11, 12, 13, 14 e 15 grupos de ligação obtidos nas saturações de 5, 10, 15 e 20 marcas em populações de retrocruzamento com 100 indivíduos.

Desconsiderando-se o nível de saturação em que não houve a recuperação de grupos de ligação, em todos os demais níveis de saturação só houve variação no número de grupos de ligação recuperados para populações com 50 indivíduos.

Na Figura 13 é apresentada a porcentagem de grupos de ligação recuperados a partir do genoma original, utilizando-se saturações de 10, 15 e 20 marcas por grupo de ligação, em populações com 50 indivíduos.

Considerando apenas as populações com 50 indivíduos e todos os níveis de saturação utilizados, representados na Figura 13, pode-se observar que, à medida que o número de marcas por grupo de ligação aumenta, o número de repetições que recuperam 10 grupos de ligação também aumenta. Na Figura 13 também pode ser observado que, mesmo aumentando a saturação, não acontece recuperação de grupos de ligação em 100% das repetições, devido ao baixo número de indivíduos nas populações, nesse caso 50 indivíduos. É constatado também nessa figura que existe diminuição no número de repetições sem formação de grupos de ligação, assim como a diminuição das repetições que apresentam número de grupos de ligação diferentes de 10. A saturação com cinco marcas por grupo de ligação não está representada na Figura 13 por não ter recuperado os 10 grupos de ligação nem apresentado diferente número desses grupos nas repetições.

Na Figura 14 é apresentada a porcentagem de grupos de ligação recuperados a partir do genoma original, utilizando-se saturações de 5, 10, 15 e 20 marcas por grupo, em populações com 100 indivíduos.

No caso das populações de retrocruzamento observadas (Figuras 5 e 6), assim como aconteceu nas populações F_2 (Figuras 3 e 4), com o aumento da saturação ou do número de indivíduos na população, independentemente um do outro, houve diminuição no número de repetições sem grupos de ligação e aumento no número de repetições que conseguem recuperar o número de grupos da população original.

Pôde ser prontamente observado que, mesmo dobrando o número de indivíduos em relação à população com 50 indivíduos, ainda não foi possível a

recuperação de grupos de ligação em 100% das repetições, mesmo na maior saturação, ou seja, com 20 marcas por grupo de ligação.

Nas populações com 100 indivíduos também parece existir a tendência à recuperação de 10 grupos de ligação nas repetições, à medida que aumenta a saturação dos grupos de ligação. Pôde ser visto que a saturação ao passar de 5 a 15 marcas por grupo de ligação, o número de repetições que recuperam 10 grupos foi de 9 para 5 marcas, 89 para 10 marcas e 95 para 15 marcas. Ao contrário do esperado para 20 marcas, apenas 93 repetições recuperam 10 grupos de ligação, contrariando a tendência observada. Apesar dessa diferença de duas repetições e de a saturação de 15 marcas ter recuperado número maior de repetições com 10 grupos de ligação, os dados da população com saturação de 20 marcas são mais confiáveis por ter o maior índice de informatividade, 2985,85 da população com 20 (0,05 cM) marcas, contra 1978,31 da população com 15 (~0,0667 cM) marcas por grupo de ligação (Quadro 15).

Nas populações com 150, 200, 400, 600 e 800 indivíduos e com as saturações de 10, 15 e 20 marcas por grupo de ligação, houve a recuperação de todos os grupos em todas as repetições, nas três saturações apresentadas.

Como comentado para populações F_2 , Weller (1986) e Patterson et al. (1988) relataram que grandes variedades de efeitos quantitativos podem ser associadas a marcadores genéticos específicos, e experimentos evidenciaram que uma porção significativa dos efeitos de características quantitativas é de uma magnitude que poderia ser prontamente detectada em experimentos com aproximadamente 1.000 indivíduos de população F_2 ou de retrocruzamento. Assim como aconteceu na populações F_2 , neste trabalho foi visto que não há necessidade de tamanho tão grande de indivíduos para que sejam detectados efeitos de características quantitativas em populações de retrocruzamento, apesar de esse tipo de população, em certos casos, necessitar de maior número de indivíduos que populações F_2 .

Lannes et al. (2004), em estudo composto por duas populações de retrocruzamento em arroz, uma com 53 e outra com 74 indivíduos e com 131 “primers” RAPD obtidos da “University of British Columbia, Inc.,” e sete que

amplificaram bandas polimórficas mais consistentes entre os pais, comentaram que mapas genéticos obtidos em seu estudo poderiam ser considerados como mapas básicos, uma vez que o número de marcadores mapeados e os tamanhos das populações de mapeamento foram menores que os utilizados em trabalhos anteriores. Uma observação interessante de seu trabalho foi a presença de sete intervalos entre marcadores que apresentaram distância de 0 cM no mapa de ligação. Segundo os referidos autores, isso pode ter acontecido em razão do pequeno tamanho da população utilizada para a formação desse mapa, que foi de 53 indivíduos. Populações com poucos genótipos podem não permitir a observação da quebra de ligação entre os marcadores e a conseqüente determinação da distância entre estes. O tamanho adequado também depende do tipo de população, sendo que populações de retrocruzamento apresentam aproximadamente metade do conteúdo informativo do que populações F_2 (LANNES et al., 2004). Ainda segundo esses autores, a grande vantagem do uso de populações F_2 é que os genótipos dessas populações são formados por dois gametas que têm seus locos em desequilíbrio de ligação, e não por um único, como no caso dos genótipos obtidos por retrocruzamento.

A identificação de distâncias de 0 cM entre marcas adjacentes também foi observada nos dados, no menor nível de saturação e nas populações de menor tamanho.

4.5.2. Correlação de Spearman entre medidas de distância

A explicação para populações F_2 também será aplicada para populações de retrocruzamento, ou seja, por se tratarem de populações hipotéticas, foi considerado não ser necessário especificar em quais repetições e grupos de ligação ocorreu inversão, sendo relatados apenas o número de repetições a recuperar 10 grupos de ligação e quantos desses grupos apresentaram inversões, independentemente do número de inversões ocorridas em cada repetição.

No Quadro 17 está apresentado o número de repetições com 10 grupos de ligação que apresentaram inversões, identificadas pela correlação de Spearman.

Quadro 17 – Nível de saturação em número de marcas por grupo de ligação, tamanho da população, número de repetições com 10 grupos de ligação e número de inversões detectadas em populações de retrocruzamento

Saturação	T.P.	Rep	NI
	50	-	-
	100	9	1
	150	31	2
5 marcas	200	35	0
	400	76	0
	600	91	0
	800	100	0
	50	64	36
	100	89	89
	150	100	0
10 marcas	200	100	0
	400	100	0
	600	100	0
	800	100	0
	50	78	76
	100	95	22
	150	100	4
15 marcas	200	100	0
	400	100	0
	600	100	0
	800	100	0
	50	84	84
	100	93	26
	150	100	11
20 marcas	200	100	3
	400	100	0
	600	100	0
	800	100	0

[†] indica que a primeira linha está vazia por falta de informações necessárias para a geração de grupos de ligação com 50 indivíduos e concentração de cinco marcas por grupo de ligação; saturação = número de marcas por grupo de ligação, do genoma simulado; T.P. = tamanho da população; Rep = número de repetições com 10 grupos de ligação; e NI = número de inversões detectadas.

Na população com 50 indivíduos e cinco marcas por grupo de ligação não houve a formação de grupos de ligação, portanto não foi possível a identificação de inversões destes.

A interpretação dos dados do Quadro 17 seguiu o critério adotado para o Quadro 6, ou seja, obtendo valores de correlação de Spearman iguais a 1 significa que as ordens das marcas nos grupos de ligação obtidos no mapeamento

das populações segregantes não foram alteradas em relação à ordem previamente conhecida do genoma original, e valor menor que 1 indica alteração na ordem das marcas em relação à ordem do genoma original.

Repetições com presença de valores de correlação de Spearman menores que 1 em um ou mais dos 10 grupos de ligação foram obtidas nas populações segregantes nos quatro níveis de saturação de genoma estudados.

No nível de saturação com cinco marcas, das nove repetições com 10 grupos de ligação a população de 100 indivíduos apresentou apenas uma repetição com valor de correlação de Spearman diferente de 1; na população com 150 indivíduos, 31 repetições recuperaram 10 grupos de ligação e duas repetições apresentaram inversão, enquanto nas demais populações, nesse grau de saturação, não foram identificadas mais inversões.

Nos outros níveis de saturação houve a recuperação de grupos de ligação na população com 50 indivíduos e na saturação de 10 marcas, em que 64 recuperaram 10 grupos de ligação e 36 apresentaram inversão, enquanto na população com 100 indivíduos 89 recuperaram 10 grupos de ligação e 89 apresentaram inversão, na saturação de 15 marcas. Na população com 50 indivíduos, 78 recuperaram 10 grupos de ligação e 76 apresentaram inversão; na população com 100 indivíduos, 95 recuperaram 10 grupos de ligação e 22 apresentaram inversão; na população com 150 indivíduos, 100 recuperaram 10 grupos de ligação e quatro apresentaram inversão na saturação de 20 marcas; nas populações com 50 indivíduos, 84 recuperaram 10 grupos de ligação e 84 apresentaram inversão; na população com 100 indivíduos, 93 recuperaram 10 grupos de ligação e 26 apresentaram inversão; na população com 150 indivíduos, 100 recuperaram 10 grupos de ligação e 11 apresentaram inversão; e, por último, na população com 200 indivíduos, 100 recuperaram 10 grupos de ligação e três apresentaram inversão.

Em todos os outros tamanhos de população, acima de 200 indivíduos, nos diferentes níveis de saturação foram recuperados os 10 grupos de ligação em 100% das repetições e não houve inversões.

Também é observado no Quadro 17 que com o aumento da saturação, em populações com 50 indivíduos, houve aumento no número de repetições que

recuperaram 10 grupos de ligação e também no número de inversões. Na população com 100 indivíduos e considerando as saturações de 10, 15 e 20 marcas, houve diminuição no número de inversões na população, comparando-se a saturação de 10 marcas com a de 15 marcas, e houve certa estabilidade quando comparadas as saturações de 15 e 20 marcas. Também houve inversão nas populações com 150 e 200 indivíduos na maior saturação. A causa do surgimento do aumento no número de inversões, apesar do aumento no número de marcas, é por ora desconhecido; mesmo assim, a confiabilidade em se usar uma população com 200 indivíduos e 20 marcas é maior que usar populações de menor tamanho e menor número de marcas, devido ao menor índice de informatividade e maior desvio-padrão do tamanho médio dos grupos de ligação.

4.5.3. Tamanho dos grupos de ligação

Para a obtenção do tamanho médio de cada grupo de ligação foi tomada a média aritmética dos tamanhos dos grupos de ligação obtidos nas repetições com 10 grupos de ligação. Os valores das médias dos tamanhos dos grupos de ligação obtidos nos sete tamanhos de população utilizados no mapeamento genético, bem como os resultados do teste de Tukey, são apresentados no Quadro 18.

O tamanho de cada grupo de ligação esperado após o mapeamento das populações segregantes é de 100 cM, uma vez que esse é o tamanho de cada grupo de ligação nos quatro níveis de saturação do genoma utilizado para geração das populações, tanto F_2 quanto de retrocruzamento.

Assim como acontece na populações F_2 , verificou-se que, a não ser na saturação de cinco marcas por grupo de ligação, nas populações de retrocruzamento não houve diferença entre as médias dos diferentes tamanhos de população dentro de cada nível de saturação.

Considerando inicialmente os grupos de ligação, no nível de saturação com cinco marcas por grupo de ligação houve diferença significativa nos diferentes tamanhos de população, nos grupos 3, 5, 6, 8, 9 e 10. Em relação às médias gerais apresentadas na última coluna do Quadro 18, as populações podem ser separadas em três grupos, que diferem entre si.

Quadro 18 – Média aritmética do tamanho de 10 grupos de ligação em sete tamanhos de populações (T.P.) de retrocruzamento em quatro níveis de saturação

Saturação	T.P.	Grupos de Ligação										Média												
		1	2	3	4	5	6	7	8	9	10													
5 marcas	50	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-			
	100	89,77 (a)	101,33 (a)	94 (ab)	89,77 (a)	96,33 (ab)	80,33 (b)	95,33 (a)	87,66 (b)	85,55 (b)	91,22 (ab)	91,13 (c)	92,45 (c)	97,08 (b)	97,35 (ab)	98,88 (a)	96,88 (a)	98,22 (ab)	98,74 (ab)	99,99 (a)	99,99 (a)	98,75 (ab)	100,33 (a)	
	150	93,66 (a)	90,79 (a)	91,43 (b)	93,57 (a)	90,65 (b)	95,54 (ab)	94,97 (a)	89,5 (b)	94,52 (ab)	89,83 (b)	92,45 (c)	97,08 (b)	97,35 (ab)	98,88 (a)	96,88 (a)	98,22 (ab)	98,74 (ab)	99,99 (a)	99,99 (a)	99,99 (a)	98,75 (ab)	100,33 (a)	
	200	95,21 (a)	91,84 (a)	98,64 (ab)	96,41 (a)	97,97 (ab)	99,55 (a)	96,88 (a)	98,22 (ab)	98,74 (ab)	97,35 (ab)	97,08 (b)	97,35 (ab)	98,88 (a)	96,88 (a)	98,22 (ab)	98,74 (ab)	99,99 (a)	99,99 (a)	99,99 (a)	99,99 (a)	98,75 (ab)	100,33 (a)	
	400	100,32 (a)	100,86 (a)	100,67 (a)	100,66 (a)	100,30 (a)	100,54 (a)	100,63 (a)	100,55 (a)	100,55 (a)	99,99 (a)	98,75 (ab)	100,33 (a)	100,65 (a)	100,33 (a)	100,63 (a)	100,55 (a)	100,55 (a)	100,55 (a)	100,55 (a)	100,55 (a)	100,55 (a)	100,65 (a)	100,65 (a)
	600	101,58 (a)	100,36 (a)	100,84 (a)	100,77 (a)	101,10 (a)	100,25 (a)	100,27 (a)	100,18 (a)	100,18 (a)	100,57 (a)	100,55 (a)	100,65 (a)	100,65 (a)	100,65 (a)	100,63 (a)	100,55 (a)	100,55 (a)	100,55 (a)	100,55 (a)	100,55 (a)	100,55 (a)	100,65 (a)	100,65 (a)
	800	100,98 (a)	100,72 (a)	100,7 (a)	101,13 (a)	100,91 (a)	101,17 (a)	101,23 (a)	100,92 (a)	100,92 (a)	100,91 (a)	100,82 (a)	100,95 (a)	100,95 (a)	100,95 (a)	100,92 (a)	100,92 (a)	100,92 (a)	100,92 (a)	100,92 (a)	100,92 (a)	100,92 (a)	100,95 (a)	100,95 (a)
10 marcas	50	102,09	99,75	102,5	102,09	105,56	100,5	102,62	102,25	102,81	102,4	102,25 (a)	102,78 (a)	102,87 (a)	102,87 (a)	102,87 (a)	102,87 (a)	102,87 (a)	102,87 (a)	102,87 (a)	102,87 (a)	102,87 (a)	102,87 (a)	
	100	102,63	104,04	102,83	103,47	102,51	103,49	102,84	100,56	103,23	102,21	102,78 (a)	102,87 (a)	102,87 (a)	102,87 (a)	102,87 (a)	102,87 (a)	102,87 (a)	102,87 (a)	102,87 (a)	102,87 (a)	102,87 (a)	102,87 (a)	
	150	102,75	104,14	102,56	103,77	102,47	102,47	101,94	103,84	102,44	102,32	102,87 (a)	102,87 (a)	102,87 (a)	102,87 (a)	102,87 (a)	102,87 (a)	102,87 (a)	102,87 (a)	102,87 (a)	102,87 (a)	102,87 (a)	102,87 (a)	
	200	103,74	102,21	103,41	103,17	103,53	102,69	102,28	102,76	103,03	102,46	102,93 (a)	102,93 (a)	102,93 (a)	102,93 (a)	102,93 (a)	102,93 (a)	102,93 (a)	102,93 (a)	102,93 (a)	102,93 (a)	102,93 (a)	102,93 (a)	
	400	102,24	101,93	101,88	103,08	102,32	102,72	102,84	102,85	102,72	102,65	102,52 (a)	102,52 (a)	102,52 (a)	102,52 (a)	102,52 (a)	102,52 (a)	102,52 (a)	102,52 (a)	102,52 (a)	102,52 (a)	102,52 (a)	102,52 (a)	
	600	103,34	102,39	103,03	102,67	102,56	102,97	102,87	102,66	103,13	102,42	102,80 (a)	102,80 (a)	102,80 (a)	102,80 (a)	102,80 (a)	102,80 (a)	102,80 (a)	102,80 (a)	102,80 (a)	102,80 (a)	102,80 (a)	102,80 (a)	
	800	102,82	102,31	103,26	103,17	102,56	102,6	102,34	102,69	103,18	102,22	102,71 (a)	102,71 (a)	102,71 (a)	102,71 (a)	102,71 (a)	102,71 (a)	102,71 (a)	102,71 (a)	102,71 (a)	102,71 (a)	102,71 (a)	102,71 (a)	
15 marcas	50	102,07	101,71	107,94	100,64	100,92	100,84	100,79	103,2	104,92	100,17	102,32 (a)	102,32 (a)	102,32 (a)	102,32 (a)	102,32 (a)	102,32 (a)	102,32 (a)	102,32 (a)	102,32 (a)	102,32 (a)	102,32 (a)	102,32 (a)	
	100	100,8	102,57	103,26	103,15	101,22	102,4	103,44	102,5	100,87	101,28	102,15 (a)	102,15 (a)	102,15 (a)	102,15 (a)	102,15 (a)	102,15 (a)	102,15 (a)	102,15 (a)	102,15 (a)	102,15 (a)	102,15 (a)		
	150	102,36	101,62	102,47	102,15	102,01	100,94	101,63	102,27	102,33	101,81	101,96 (a)	101,96 (a)	101,96 (a)	101,96 (a)	101,96 (a)	101,96 (a)	101,96 (a)	101,96 (a)	101,96 (a)	101,96 (a)	101,96 (a)	101,96 (a)	
	200	102,9	101,09	103,62	102,34	103,74	102,17	103,14	102,3	102,97	103,17	102,74 (a)	102,74 (a)	102,74 (a)	102,74 (a)	102,74 (a)	102,74 (a)	102,74 (a)	102,74 (a)	102,74 (a)	102,74 (a)	102,74 (a)	102,74 (a)	
	400	102,63	103,22	102,74	102,88	103,01	101,95	101,24	101,46	102,17	102,5	102,38 (a)	102,38 (a)	102,38 (a)	102,38 (a)	102,38 (a)	102,38 (a)	102,38 (a)	102,38 (a)	102,38 (a)	102,38 (a)	102,38 (a)	102,38 (a)	
	600	102,17	101,94	101,92	102,54	101,79	102,07	102	102,37	101,97	101,35	101,99 (a)	101,99 (a)	101,99 (a)	101,99 (a)	101,99 (a)	101,99 (a)	101,99 (a)	101,99 (a)	101,99 (a)	101,99 (a)	101,99 (a)	101,99 (a)	
	800	102,31	101,86	102,25	101,94	101,83	102,22	102,87	102,64	102,52	102,2	102,26 (a)	102,26 (a)	102,26 (a)	102,26 (a)	102,26 (a)	102,26 (a)	102,26 (a)	102,26 (a)	102,26 (a)	102,26 (a)	102,26 (a)	102,26 (a)	
20 marcas	50	104,26	101,21	105,42	100,59	103,3	103,02	102,76	105,19	103,8	102,54	103,21 (a)	103,21 (a)	103,21 (a)	103,21 (a)	103,21 (a)	103,21 (a)	103,21 (a)	103,21 (a)	103,21 (a)	103,21 (a)	103,21 (a)	103,21 (a)	
	100	102,35	101,9	102,56	101,94	103,19	102,69	101,81	101,29	100,04	101,38	101,92 (a)	101,92 (a)	101,92 (a)	101,92 (a)	101,92 (a)	101,92 (a)	101,92 (a)	101,92 (a)	101,92 (a)	101,92 (a)	101,92 (a)	101,92 (a)	
	150	100,87	101,13	102,98	101,89	100,31	101,84	100,89	101,23	101,82	99,98	101,25 (a)	101,25 (a)	101,25 (a)	101,25 (a)	101,25 (a)	101,25 (a)	101,25 (a)	101,25 (a)	101,25 (a)	101,25 (a)	101,25 (a)	101,25 (a)	
	200	101,24	102,07	102,85	102,15	101,39	102,66	102,85	103,14	101,11	101,78	102,12 (a)	102,12 (a)	102,12 (a)	102,12 (a)	102,12 (a)	102,12 (a)	102,12 (a)	102,12 (a)	102,12 (a)	102,12 (a)	102,12 (a)	102,12 (a)	
	400	101,29	101,3	102,66	101,89	102,23	101,37	101,19	101,53	100,99	101,38	101,58 (a)	101,58 (a)	101,58 (a)	101,58 (a)	101,58 (a)	101,58 (a)	101,58 (a)	101,58 (a)	101,58 (a)	101,58 (a)	101,58 (a)	101,58 (a)	
	600	101,65	101,17	101,37	101,61	100,85	101,54	101,21	100,77	101,55	102,36	101,41 (a)	101,41 (a)	101,41 (a)	101,41 (a)	101,41 (a)	101,41 (a)	101,41 (a)	101,41 (a)	101,41 (a)	101,41 (a)	101,41 (a)	101,41 (a)	
	800	101,73	101,89	101,38	102,14	101,77	102,05	102,15	101,96	101,37	101,79	101,82 (a)	101,82 (a)	101,82 (a)	101,82 (a)	101,82 (a)	101,82 (a)	101,82 (a)	101,82 (a)	101,82 (a)	101,82 (a)	101,82 (a)	101,82 (a)	

¹ indica que a primeira linha está vazia por falta de informações necessárias para a geração de grupos de ligação com 50 indivíduos e concentração de cinco marcas por grupo de ligação; e ² indica que dentro dos parêntesis estão as letras correspondentes ao resultado da análise de médias feita pelo teste de Tukey no nível de significância de 1%. Nas colunas, médias seguidas de uma mesma letra não diferem entre si, pelo teste de Tukey.

Pela observação das médias gerais, ainda no menor nível de saturação é constatado que, à medida que o tamanho da população aumenta, o tamanho dos grupos de ligação também aumenta até se aproximar do tamanho originalmente utilizado para simulação das populações, que é de 100 cM.

Nos demais níveis de saturação avaliados não houve variação significativa no tamanho dos grupos de ligação, nem das médias gerais dos grupos de ligação com o aumento no número de indivíduos nas populações.

Apesar da estabilidade no tamanho médio dos grupos de ligação, à medida que aumenta o tamanho da população segregante, a amplitude de variação dos tamanhos de grupos de ligação obtidos entre repetições diminui. Essa redução na amplitude de variação dos tamanhos de grupos de ligação pode ser observada pela redução no desvio-padrão com o aumento no número de indivíduos nas populações segregantes. Pode-se ter como exemplo as médias gerais de desvio-padrão para menor e maior nível de saturação, em que o desvio-padrão foi de 13,62 para 3,19 cM no menor nível de saturação, na população com 150 e 800 indivíduos, respectivamente, e variou de 15,03 a 3,53 cM nas populações com 50 e 800 indivíduos, respectivamente, no maior nível de saturação (Quadro 19).

No Quadro 19, observando os valores de desvio-padrão na última coluna em todos os níveis de saturação, somente no nível de saturação de cinco marcas o desvio para uma população com 150 indivíduos é maior que em uma população com 100 indivíduos, sendo 13,62 e 11,84 cM, respectivamente. Nos demais tamanhos de população, a diminuição no valor de desvio se mantém com o aumento no tamanho da população, passando de 9,80 cM na população com 200 indivíduos para 3,19 cM naquela com 800 indivíduos. Considerando os demais níveis de saturação, o desvio-padrão variou de 13,54 a 3,42 cM na saturação com 10 marcas, 13,91 a 3,42 cM na saturação com 15 marcas e 15,03 a 3,53 cM na saturação com 20 marcas, sendo os de maiores desvios referentes às populações com 50 indivíduos e os valores de menores desvios, referentes às populações com 800 indivíduos.

Quadro 19 – Desvio-padrão do tamanho de 10 grupos de ligação em sete tamanhos de populações (T.P.) de retrocruzamento em quatro níveis de saturação

Saturação	T.P.	Grupos de Ligação										Média	
		1	2	3	4	5	6	7	8	9	10		
5 marcas	50	-	-	-	-	-	-	-	-	-	-	-	¹
	100	11,98	6,42	9,39	17,59	5,67	18,12	8,84	12,66	15,64	12,12	11,84	
	150	12,12	14,1	12,68	12,97	15,64	12,58	14,19	16,18	12,32	13,47	13,62	
	200	13,25	12,5	6,68	9,07	10,16	8,67	10,13	10,14	7,11	10,38	9,80	
	400	5,09	4,14	5,48	4,41	5,09	4,88	4,81	5,36	6,09	9,43	5,47	
	600	3,43	4,25	4,3	4,93	3,58	3,36	3,97	4,16	3,41	3,96	3,93	
	800	3,4	3,02	3,15	2,98	3,18	2,88	3,14	2,87	3,38	3,91	3,19	
10 marcas	50	15,17	15,97	14,14	12,06	13,2	13,66	12,56	14,77	11,76	12,14	13,54	
	100	9,78	10,05	9,3	9,42	9,56	8,89	9,9	9,16	10,17	9,83	9,60	
	150	7,85	7,26	8,27	6,66	8,36	7,42	7,49	8,13	7,55	8,25	7,72	
	200	6,63	7,68	5,9	6,67	6,47	6,02	6,83	7,21	6,73	6,59	6,67	
	400	4,96	5,2	5,09	4,61	4,95	4,22	5,13	4,64	4,28	4,45	4,75	
	600	3,83	4,11	3,77	3,76	3,61	3,21	3,44	3,87	3,99	3,99	3,75	
	800	3,37	3,39	3,55	3,51	3,45	3,49	3,84	3,62	3,15	2,87	3,42	
15 marcas	50	14,45	11,05	13,96	13,17	14,42	15,56	14,41	15,9	13,75	12,52	13,91	
	100	9,7	9,91	10,36	9,7	11,45	10,56	9,68	9,31	9,8	9,78	10,02	
	150	7,62	7,57	7,21	8,37	6,76	7,47	8,36	9,71	7,58	8,87	7,95	
	200	7,35	7,32	6,76	7,07	7,68	7,49	7,58	6,57	7,2	6,03	7,10	
	400	5,75	4,82	4,76	4,5	5,35	4,69	5,69	4,55	4,84	4,7	4,96	
	600	3,69	3,85	4,38	4	3,89	3,44	3,79	3,98	3,97	3,96	3,89	
	800	2,93	3,8	3,05	3,32	3,68	3,99	3,29	3,53	3,35	3,28	3,42	
20 marcas	50	13,79	15,07	15,58	13,78	16,58	16,67	13,71	14,02	13,48	17,71	15,03	
	100	11,91	10,32	8,76	9,82	10,72	10,79	9,34	10,29	10,16	10,15	10,22	
	150	8,79	8,91	8,66	8,51	8,21	7,63	8,17	7,93	8,81	7,89	8,35	
	200	7,91	6,63	6,65	6,96	7,77	7,94	7,2	7,35	6,43	6,66	7,15	
	400	5,39	4,45	5,08	5,12	5,2	5,02	4,57	4,38	5,14	5,54	4,98	
	600	4,51	3,97	4,25	4,2	4,22	4,3	3,84	3,86	4,15	3,81	4,11	
	800	4,17	3,46	3,64	3,7	3,13	3,25	3,8	3,24	3,08	3,9	3,53	

¹ indica que a primeira linha está vazia por falta de informações necessárias para a geração de grupos de ligação com 50 indivíduos e concentração de cinco marcas por grupo de ligação.

A tendência de redução da amplitude de variação observada nas médias do tamanho dos grupos de ligação constatada pela observação dos desvios acontece também entre os grupos de ligação dos diferentes tamanhos de população, a não ser na maioria dos grupos de ligação no nível de saturação com cinco marcas por grupo de ligação, como apresentado no Quadro 19. Nos demais grupos de ligação e nos outros níveis de saturação, apesar de em alguns casos os desvios serem um pouco maiores em populações com maior número de indivíduos, as diferenças são pequenas, chegando a décimos ou centésimos de unidade, como pode ser

observado no grupo de ligação 4 no nível de saturação de 10 marcas, nas populações com 150 e 200 indivíduos, 6,66 e 6,67, respectivamente. Considerando as populações de tamanho a partir de 200 indivíduos em todos os níveis de saturação, em relação à média geral dos desvios, os valores foram maiores em relação aos desvios das populações F_2 analisados. Nas populações de retrocruzamento, esses valores foram menores que 8 cM, enquanto nas populações F_2 , menores que 5. É importante observar que, considerando os três maiores níveis de saturação, de acordo com o representado na última coluna, a média dos desvios-padrão é menor que 5, a não ser nas populações com nível de saturação de cinco marcas por grupo de ligação.

Pela observação das médias gerais, ainda no menor nível de saturação é constatado que, à medida que o tamanho da população aumenta, o tamanho dos grupos de ligação também aumenta até se aproximar do tamanho originalmente utilizado para simulação das populações, que é de 100 cM.

4.5.4. Média das distâncias entre marcas adjacentes

A distância média de marcas adjacentes ao longo de cada grupo de ligação foi obtida fazendo-se duas médias aritméticas sucessivas. A primeira média foi feita apenas entre os valores de distância encontrados dentro de cada grupo de ligação. A segunda foi obtida pela média aritmética das médias anteriormente obtidas. Nesse processo só foram utilizadas repetições que tinham 10 grupos de ligação, sendo as demais desconsideradas.

Os valores de distância média entre marcas adjacentes e seu desvio-padrão obtido nos vários tamanhos de populações segregantes utilizados no mapeamento genético estão apresentados nos Quadros 20 e 21, respectivamente.

Visscher et al. (1996), utilizando técnicas de simulação, avaliaram a eficiência da introgressão de genes em populações de retrocruzamento derivadas de linhas endogâmicas. Após a obtenção dos dados, ficou constatado que marcadores foram eficientes tanto para introgressão de alelos quanto para seleção de genótipos receptores, simultaneamente. Ao serem consideradas marcas espaçadas de 10 a 20 cM, houve vantagem de uma a duas gerações de seleção de retrocruzamentos, em relação ao tempo gasto no melhoramento clássico, quanto à seleção fenotípica.

Quadro 20 – Distância média entre marcas adjacentes nas repetições que recuperaram 10 grupos de ligação, em populações de retrocruzamentos, avaliados em quatro níveis de saturação

Saturação	T.P.	Grupos de Ligação										Média	
		1	2	3	4	5	6	7	8	9	10		
5 marcas	50	-	-	-	-	-	-	-	-	-	-	-	¹
	100	23,79	25,33	24,21	24,30	24,08	24,38	24,55	24,70	24,05	24,25	24,36 (b) ²	
	150	24,39	24,73	23,97	24,58	24,20	24,65	25,15	24,41	24,85	24,69	24,56 (b)	
	200	25,05	24,37	24,83	24,84	25,19	25,23	24,76	24,87	24,87	25,03	24,90 (ab)	
	400	25,25	25,21	25,25	25,25	25,16	25,21	25,23	25,22	25,15	25,24	25,22 (a)	
	600	25,39	25,16	25,27	25,33	25,27	25,19	25,13	25,11	25,14	25,20	25,22 (a)	
	800	25,24	25,18	25,17	25,28	25,22	25,29	25,30	25,23	25,22	25,26	25,24 (a)	
10 marcas	50	11,34	11,17	11,33	11,35	11,75	11,16	11,44	11,31	11,42	11,37	11,36 (a)	
	100	11,40	11,56	11,42	11,49	11,39	11,49	11,42	11,17	11,47	11,35	11,42 (a)	
	150	11,41	11,57	11,39	11,53	11,38	11,38	11,32	11,53	11,38	11,36	11,43 (a)	
	200	11,52	11,35	11,48	11,46	11,50	11,41	11,36	11,41	11,44	11,38	11,43 (a)	
	400	11,36	11,32	11,32	11,45	11,36	11,41	11,42	11,42	11,41	11,40	11,39 (a)	
	600	11,48	11,37	11,44	11,40	11,39	11,44	11,43	11,40	11,45	11,38	11,42 (a)	
	800	11,42	11,36	11,47	11,46	11,39	11,39	11,37	11,41	11,46	11,35	11,41 (a)	
15 marcas	50	7,29	7,26	7,71	7,18	7,20	7,20	7,19	7,37	7,49	7,15	7,30 (a)	
	100	7,19	7,32	7,375	7,36	7,23	7,31	7,38	7,32	7,20	7,23	7,29 (a)	
	150	7,31	7,25	7,31	7,29	7,28	7,21	7,25	7,30	7,30	7,27	7,28 (a)	
	200	7,35	7,22	7,40	7,31	7,41	7,29	7,36	7,30	7,35	7,36	7,33 (a)	
	400	7,33	7,37	7,33	7,34	7,35	7,28	7,23	7,24	7,29	7,32	7,31 (a)	
	600	7,29	7,28	7,27	7,32	7,27	7,29	7,28	7,31	7,27	7,23	7,28 (a)	
	800	7,30	7,27	7,30	7,28	7,27	7,30	7,34	7,33	7,32	7,30	7,30 (a)	
20 marcas	50	5,48	5,33	5,54	5,29	5,43	5,42	5,40	5,53	5,46	5,39	5,43 (a)	
	100	5,38	5,36	5,39	5,36	5,43	5,40	5,35	5,33	5,26	5,33	5,36 (a)	
	150	5,30	5,32	5,41	5,34	5,27	5,36	5,31	5,32	5,35	5,26	5,32 (a)	
	200	5,32	5,37	5,41	5,37	5,33	5,40	5,41	5,42	5,32	5,35	5,37 (a)	
	400	5,33	5,33	5,40	5,36	5,38	5,33	5,32	5,34	5,31	5,33	5,34 (a)	
	600	5,35	5,32	5,33	5,34	5,30	5,34	5,32	5,30	5,34	5,38	5,33 (a)	
	800	5,35	5,36	5,33	5,37	5,35	5,37	5,37	5,36	5,33	5,35	5,35 (a)	

¹ indica que a primeira linha está vazia por falta de informações necessárias para a geração de grupos de ligação com 50 indivíduos e concentração de cinco marcas por grupo de ligação; ² indica que dentro dos parêntesis estão as letras correspondentes ao resultado da análise de médias feita pelo teste de Tukey no nível de significância de 1%.

Quadro 21 – Desvio-padrão das médias das distâncias entre marcas adjacentes nas repetições que recuperaram 10 grupos de ligação, em populações de retrocruzamento, avaliados em quatro níveis de saturação

Saturação	T.P.	Grupos de Ligação										
		1	2	3	4	5	6	7	8	9	10	
5 marcas	50	-	-	-	-	-	-	-	-	-	-	¹
	100	1,9	1,6	1,8	1,81	1,41	1,88	1,59	1,22	2,7	2,58	
	150	2,0	1,68	1,71	1,43	1,58	1,5	1,57	1,79	1,28	1,36	
	200	1,6	1,37	1,19	1,63	1,32	1,4	1,21	1,77	1,57	1,43	
	400	0,98	1,03	1,14	0,98	1,14	0,95	0,91	1,17	1,02	1,15	
	600	0,85	0,85	0,81	0,83	0,89	0,77	0,73	0,8	0,85	0,74	
	800	0,85	0,75	0,78	0,74	0,79	0,72	0,78	0,71	0,84	0,74	
10 marcas	50	1,68	1,74	1,56	1,30	1,47	1,51	1,35	1,58	1,30	1,34	
	100	1,08	1,11	1,03	1,04	1,06	0,98	1,10	1,01	1,13	1,09	
	150	0,87	0,80	0,91	0,74	0,92	0,82	0,83	0,90	0,83	0,91	
	200	0,73	0,85	0,65	0,74	0,71	0,66	0,75	0,80	0,74	0,73	
	400	0,55	0,57	0,56	0,51	0,55	0,46	0,57	0,51	0,47	0,49	
	600	0,42	0,45	0,41	0,41	0,40	0,35	0,38	0,43	0,44	0,44	
	800	0,37	0,37	0,39	0,39	0,38	0,38	0,42	0,40	0,35	0,31	
15 marcas	50	1,03	0,78	0,99	0,94	1,03	1,11	1,02	1,13	0,98	0,89	
	100	0,69	0,70	0,74	0,69	0,81	0,75	0,69	0,66	0,70	0,69	
	150	0,54	0,54	0,51	0,59	0,48	0,53	0,59	0,69	0,54	0,63	
	200	0,52	0,52	0,48	0,50	0,54	0,53	0,54	0,46	0,51	0,43	
	400	0,41	0,34	0,34	0,32	0,38	0,33	0,40	0,32	0,34	0,33	
	600	0,26	0,27	0,31	0,28	0,27	0,24	0,27	0,28	0,28	0,28	
	800	0,2	0,27	0,21	0,23	0,26	0,28	0,23	0,25	0,23	0,23	
20 marcas	50	0,72	0,79	0,82	0,72	0,87	0,87	0,72	0,73	0,70	0,93	
	100	0,62	0,54	0,46	0,51	0,56	0,56	0,49	0,54	0,53	0,53	
	150	0,46	0,46	0,45	0,44	0,43	0,40	0,43	0,41	0,46	0,41	
	200	0,41	0,34	0,35	0,36	0,40	0,41	0,37	0,38	0,33	0,35	
	400	0,28	0,23	0,26	0,26	0,27	0,26	0,24	0,23	0,27	0,29	
	600	0,23	0,20	0,22	0,22	0,22	0,22	0,20	0,20	0,21	0,20	
	800	0,21	0,18	0,19	0,19	0,16	0,17	0,20	0,17	0,16	0,20	

¹ indica que a primeira linha está vazia por falta de informações necessárias para a geração de grupos de ligação com 50 indivíduos e concentração de cinco marcas por grupo de ligação.

As populações de retrocruzamento, assim como as populações F₂ já comentadas, foram simuladas a partir de genótipos já conhecidos, e as distâncias médias entre marcas adjacentes esperadas após o mapeamento das populações segregantes são de 20, 10, 6,67 e 5 cM, nas populações segregantes geradas a partir dos genomas com níveis de saturação de 5, 10, 15 e 20 marcas, respectivamente. Como pode ser observado no Quadro 20, de forma similar ao

ocorrido na populações F_2 , independentemente do tamanho da população ou do nível de saturação utilizado, todos os valores ficaram acima dos inicialmente esperados.

Considerando o teste de médias feito com $P < 0,01$, foi encontrado diferença apenas no menor nível de saturação; nos demais níveis, independentemente do tamanho das populações, não foram encontradas diferenças significativas entre os valores das médias.

Analisando o nível de saturação com cinco marcas, apesar de não diferir das populações com 200 indivíduos, as populações com 100 e 150 diferem das maiores populações, com 400, 600 e 800 indivíduos. Nos três maiores níveis de saturação não foi encontrado diferença significativa nas médias das distâncias entre marcas adjacentes nos diferentes tamanhos de população. As letras nas médias dos três maiores níveis de saturação representam o que também acontece em relação aos grupos de ligação.

A falta de tendência à diminuição do tamanho da distância entre marcas adjacentes pode ser claramente vista, por exemplo, no grupo de ligação 4, na saturação com 20 marcas por grupo de ligação, em que na população com 50 indivíduos a distância entre marcas é 5,29, enquanto na população com 800 indivíduos a distância entre médias é 5,37; como visto, ocorreu pequeno aumento nos valores. Percebeu-se que, nesse mesmo grupo de ligação, existe pequena variação nos valores, mas isso não significa tendência de diminuição. Essa variação também pode ser observada nos demais grupos de ligação, em todos os níveis de saturação.

A diferença encontrada, que poder ser observada no Quadro 20, é que, à medida que aumenta o nível de saturação, os valores das médias diminuem independentemente do tamanho das populações. Embora a diminuição nos valores das médias entre marcas adjacentes ocorra apenas em função do aumento do nível de saturação, o desvio-padrão apresenta comportamento diferente.

No Quadro 21 estão os desvios das médias das distâncias entre marcas adjacentes nas repetições que recuperaram 10 grupos de ligação, avaliados em quatro níveis de saturação, em populações de retrocruzamento. Observou-se que,

apesar dos valores muito próximos, pelo desvio-padrão pode ser verificado que existe tendência à diminuição da amplitude de variação nos valores das médias das distâncias entre marcas adjacentes, de modo que, quanto maior a população, menor o desvio-padrão. No nível de saturação com cinco marcas por grupo de ligação, essa tendência pode ser mostrada nas populações iguais e superiores a 400 indivíduos, à exceção do grupo de ligação 7, no qual o valor do desvio-padrão da população com 800 indivíduos (0,78) foi maior que o do desvio-padrão da população com 600 indivíduos. Nas populações com 100, 150 e 200 indivíduos, essa tendência existiu apenas nos grupos de ligação 3, 6 e 7.

Nos níveis de saturação com 10 e 15 marcas por grupo de ligação, apenas três grupos apresentaram valores maiores em populações com maior número de indivíduos, como nas populações com 150 e 200 indivíduos no grupo de ligação 5, no nível de saturação com 15 marcas.

Lannes et al. (2004) construíram mapas genéticos de retrocruzamento, para análise de regeneração de anteras de arroz irrigado em vidro e mapeamento de QTL. Primeiramente foi construído um esqueleto dos grupos de ligação utilizando distância máxima de 25 cM, com as distâncias no mapa determinadas com o emprego da função de Kosambi. Após o estabelecimento do esqueleto dos grupos de ligação, foram acrescentados mais marcadores aos mapas sob distância máxima de 30 cM. Com isso foi detectado um QTL que explicou 26,7% da variação total da característica na população segregante. Em estudo de simulação para demonstração do método, concluiu-se que o poder adequado para uma variedade de modelos de densidade de marcadores e efeitos foi alcançado. Porém, os mapas genéticos obtidos, segundo esses autores, puderam ser considerados mapas básicos, uma vez que o número de marcadores mapeados e o tamanho das populações de mapeamento foram menores que em trabalhos anteriores desses autores.

4.5.5. Variância das distâncias entre marcas adjacentes

A partir das distâncias entre as marcas adjacentes obtidas nos grupos de ligação foi feita a estimação da variância amostral, como apresentado no Quadro 22. Como em quadros anteriores, para cada grupo de ligação são dadas as médias aritméticas, nesse caso, das variâncias obtidas em cada repetição, em que houve a formação de 10 grupos de ligação no mapeamento das populações de retrocruzamento simulado.

Os valores de variância encontrados na análise dos mapas obtidos das populações segregantes simulados são referentes aos erros para qualquer tamanho de população e nível de saturação dos genomas avaliados, uma vez que os genomas utilizados na geração das populações segregantes tinham seus marcadores distribuídos de forma equidistante dentro dos grupos de ligação em todos os níveis de saturação, sendo que no genoma com nível de saturação de cinco marcas por grupo de ligação os marcadores encontravam-se distribuídos a distâncias equidistantes de 20 cM ao longo dos grupos de ligação. No genoma com nível de saturação de 10 marcas por grupo de ligação, os marcadores se achavam distribuídos à distância de 10 cM ao longo dos grupos de ligação. No genoma com nível de saturação de 15 marcas por grupo de ligação, os marcadores estavam distribuídos à distância de cerca de 6,67 cM e, finalmente, no genoma com nível de saturação de 20 marcas por grupo de ligação os marcadores encontravam-se distribuídos à distância de 5 cM ao longo dos grupos de ligação.

Como visto anteriormente, os valores de variância podem ser interpretados de forma que, quanto menores os valores de variância, mais equidistantes estarão distribuídas as marcas dentro dos grupos de ligação e, conseqüentemente, menor o erro. Portanto, quanto menores os valores de variância, mais próximos estarão os valores do esperado em cada nível de saturação, indicando uma boa recuperação do genoma com o mapeamento das populações segregantes.

Quadro 22 – Variância das distâncias entre médias adjacentes nas repetições que recuperaram 10 grupos de ligação, em populações de retrocruzamento, avaliados em quatro níveis de saturação

Saturação	T.P.	Grupos de Ligação										Médias		
		1	2	3	4	5	6	7	8	9	10			
5 marcas	50	-	-	-	-	-	-	-	-	-	-	-	-	-
	100	8,18 (ab)	14,72 (a)	13,76 (a)	6,67 (ab)	11,12 (a)	21,31 (a)	15,20 (a)	17,03 (a)	10,46 (a)	11,47 (a)	12,99 (a)	9,52 (b)	
	150	9,73 (a)	8,51 (ab)	11,53 (a)	9,26 (a)	9,59 (a)	9,97 (b)	7,07 (ab)	11,28 (a)	9,29 (a)	8,95 (a)	7,22 (c)	4,36 (d)	
	200	8,49 (a)	7,77 (ab)	6,42 (ab)	6,19 (ab)	9,06 (a)	6,34 (bc)	7,41 (ab)	6,51 (ab)	7,4 (a)	6,57 (ab)	3,12 (de)	2,27 (e)	
	400	4,59 (ab)	3,75 (bc)	4,63 (b)	4,56 (ab)	4,39 (a)	4,01 (bc)	4,58 (b)	4,64 (b)	4,01 (a)	4,41 (ab)	3,12 (de)	2,27 (e)	
	600	3,23 (ab)	2,92 (c)	3,15 (b)	3,32 (b)	3,03 (a)	2,98 (c)	2,85 (b)	2,66 (b)	3,45 (a)	3,59 (ab)	3,12 (de)	2,27 (e)	
10 marcas	800	2,14 (b)	2,51 (c)	2,46 (b)	2,19 (b)	2,47 (a)	1,87 (c)	2,29 (b)	2,31 (b)	2,22 (a)	2,20 (b)	2,27 (e)	2,27 (e)	
	50	19,32 (a)	19,26 (a)	18,62 (a)	20,04 (a)	18,54 (a)	18,36 (a)	20,36 (a)	21,71 (a)	18,57 (a)	18,88 (a)	19,37 (a)	10,36 (b)	
	100	11,35 (b)	10,56 (b)	10,75 (b)	11,17 (b)	9,61 (b)	11,56 (b)	9,99 (b)	9,71 (b)	10,11 (b)	8,81 (b)	6,56 (c)	6,56 (c)	
	150	5,98 (c)	7,18 (c)	6,07 (c)	6,74 (c)	6,97 (bc)	6,73 (c)	6,43 (bc)	6,36 (c)	6,61 (c)	6,47 (bc)	5,04 (d)	2,49 (e)	
	200	5,70 (c)	4,99 (cd)	5,19 (cd)	5,33 (cd)	4,62 (cd)	4,81 (cd)	5,04 (cd)	5,36 (cd)	4,77 (cd)	4,59 (cd)	2,48 (de)	1,71 (ef)	
	400	2,33 (d)	2,57 (de)	2,43 (cde)	2,65 (de)	2,33 (de)	2,53 (de)	2,77 (de)	2,40 (de)	2,37 (de)	2,37 (de)	1,76 (de)	1,27 (f)	
15 marcas	600	1,77 (d)	1,63 (e)	1,63 (de)	1,80 (e)	1,63 (e)	1,65 (e)	1,74 (de)	1,87 (e)	1,65 (de)	1,76 (de)	1,71 (ef)	1,27 (f)	
	800	1,35 (d)	1,26 (e)	1,25 (e)	1,20 (e)	1,31 (e)	1,33 (e)	1,25 (e)	1,17 (e)	1,3 (e)	1,24 (e)	1,27 (f)	1,27 (f)	
	50	13,91 (a)	13,56 (a)	14,53 (a)	13,90 (a)	13,47 (a)	12,71 (a)	15,26 (a)	14,07 (a)	15,04 (a)	13,74 (a)	14,02 (a)	6,80 (b)	
	100	6,93 (b)	7,00 (b)	7,02 (b)	6,60 (b)	6,97 (b)	6,77 (b)	6,50 (b)	6,78 (b)	6,75 (b)	6,66 (b)	4,46 (c)	4,46 (c)	
	150	4,55 (c)	4,38 (c)	4,53 (bc)	4,44 (c)	4,54 (bc)	4,29 (c)	4,33 (bc)	4,43 (bc)	4,60 (bc)	4,48 (c)	3,40 (d)	3,40 (d)	
	200	3,44 (cd)	3,35 (cd)	3,60 (cd)	3,36 (cd)	3,48 (cd)	3,36 (cd)	3,35 (cd)	3,46 (cd)	3,27 (cd)	3,36 (cd)	1,65 (e)	1,12 (ef)	
20 marcas	400	1,70 (de)	1,71 (de)	1,71 (de)	1,61 (de)	1,67 (de)	1,67 (de)	1,69 (de)	1,49 (cd)	1,59 (de)	1,70 (de)	1,07 (e)	1,12 (ef)	
	600	1,07 (e)	1,09 (e)	1,14 (de)	1,19 (e)	1,12 (e)	1,10 (e)	1,11 (de)	1,17 (d)	1,19 (de)	1,17 (d)	0,85 (f)	0,85 (f)	
	800	0,89 (e)	0,87 (e)	0,77 (e)	0,90 (e)	0,87 (e)	0,87 (e)	0,83 (e)	0,83 (d)	0,85 (e)	0,84 (e)	0,85 (f)	0,85 (f)	
	50	10,79 (a)	10,67 (a)	11,21 (a)	10,49 (a)	12,40 (a)	10,69 (a)	11,05 (a)	11,71 (a)	11,55 (a)	10,38 (a)	11,09 (a)	5,22 (b)	
	100	5,06 (b)	4,73 (b)	5,36 (b)	5,63 (b)	4,97 (b)	5,66 (b)	5,09 (b)	5,35 (b)	5,04 (b)	5,32 (b)	3,35 (c)	3,35 (c)	
	150	3,24 (c)	3,29 (bc)	3,61 (c)	3,20 (c)	3,15 (bc)	3,59 (bc)	3,33 (bc)	3,26 (c)	3,77 (bc)	3,06 (c)	2,52 (d)	2,52 (d)	
20 marcas	200	2,67 (cd)	2,54 (bcd)	2,61 (cd)	2,59 (cd)	2,54 (bc)	2,40 (cd)	2,43 (cd)	2,56 (cd)	2,41 (cd)	2,47 (cd)	1,28 (e)	1,28 (e)	
	400	1,24 (de)	1,24 (cd)	1,35 (de)	1,25 (de)	1,34 (c)	1,22 (d)	1,22 (d)	1,31 (cd)	1,28 (d)	1,35 (de)	0,83 (ef)	0,83 (ef)	
	600	0,81 (e)	0,85 (d)	0,80 (e)	0,83 (e)	0,80 (c)	0,84 (d)	0,82 (d)	0,85 (d)	0,84 (d)	0,83 (e)	0,63 (f)	0,63 (f)	
	800	0,62 (e)	0,64 (d)	0,63 (e)	0,62 (e)	0,66 (c)	0,62 (d)	0,63 (d)	0,62 (d)	0,62 (d)	0,61 (e)	0,63 (f)	0,63 (f)	

¹ indica que a primeira linha está vazia por falta de informações necessárias para a geração de grupos de ligação com 50 indivíduos e concentração de cinco marcas por grupo de ligação; ² indica que dentro dos parêntesis estão as letras correspondentes ao resultado da análise de médias feita pelo teste de Tukey no nível de significância de 1%. Nas colunas, médias seguidas de uma mesma letra não diferem entre si, pelo teste de Tukey.

Analisando o efeito do tamanho de população, dentro do nível de saturação de genoma de cinco marcas por grupo de ligação verifica-se redução na variância média das distâncias entre marcas adjacentes, a não ser nos grupos de ligação 1, 4 e 7, à medida que aumenta o tamanho da população segregante. Um exemplo da redução na variância, no grupo 10 tem-se que o valor da variância média foi de 11,47 e 2,20, nos tamanhos de população de 100 e 800 indivíduos, respectivamente (Quadro 22).

Além da redução na variância média, observa-se redução na amplitude de variação dos valores de variância entre repetições, à medida que aumenta o tamanho da população segregante. Essa redução pode ser demonstrada pela redução no desvio-padrão. Como exemplo, podem-se observar os desvios no grupo de ligação 10 das populações citadas anteriormente, sendo que o desvio-padrão foi de 8,64 e 1,43, respectivamente, nos tamanhos de população de 100 e 800 indivíduos (Quadro 23).

Na análise da média geral houve diferença significativa entre médias dos diferentes tamanhos de população, sendo que não diferiram entre si as médias das populações de tamanhos 400 e 600 e aquelas de tamanhos 600 e 800 (Quadro 22).

Nos sete tamanhos de população segregantes analisados dentro do nível de saturação de 10 marcas por grupo de ligação, assim como no nível de saturação anterior, houve redução na variância média das distâncias entre marcas adjacentes, à medida que aumentou o tamanho da população segregante; nesse caso, em todos os grupos de ligação. Como exemplo, mantendo o grupo de ligação 10 como referência, as variâncias foram 18,88 e 1,24, nas populações com 50 e 800 indivíduos, respectivamente (Quadro 22). Em relação ao desvio-padrão, os valores foram 5,57 e 0,34, nessas mesmas populações (Quadro 23). Também é constatada uma redução na amplitude de variação dos valores de variância entre repetições, com o aumento no tamanho da população segregante. Essa redução é observável pelo desvio-padrão (Quadro 23).

Quadro 23 – Desvio-padrão das médias das variâncias das distâncias entre marcas adjacentes nas repetições que recuperaram 10 grupos de ligação, em populações de retrocruzamento, avaliados em quatro níveis de saturação

Saturação	T.P.	Grupos de ligação										
		1	2	3	4	5	6	7	8	9	10	
5 marcas	50	-	-	-	-	-	-	-	-	-	-	-
	100	7,06	7,56	5,93	7,62	8,87	6,39	6,48	8,29	5,28	8,64	
	150	5,47	5,06	4,19	4,58	4,81	5,38	7,50	5,84	5,35	4,49	
	200	3,93	3,86	3,28	3,51	4,16	4,76	4,00	4,07	3,39	3,27	
	400	2,29	2,65	2,61	2,88	2,91	2,20	2,36	2,70	2,29	2,15	
	600	1,67	1,71	1,57	1,82	1,76	2,13	1,67	1,47	1,92	2,01	
	800	1,42	1,36	1,45	1,23	1,11	1,47	1,30	1,27	1,38	1,43	
10 marcas	50	4,81	5,69	6,54	5,61	4,94	5,45	5,19	7,52	5,96	5,57	
	100	3,22	2,61	2,80	2,81	3,04	3,22	2,92	4,11	2,91	3,09	
	150	1,88	1,74	1,85	1,92	1,72	2,01	2,21	1,86	1,84	1,93	
	200	1,48	1,29	1,18	1,32	1,55	1,50	1,36	1,44	1,50	1,51	
	400	0,65	0,62	0,74	0,72	0,73	0,91	0,70	0,69	0,84	0,78	
	600	0,52	0,51	0,48	0,45	0,44	0,51	0,46	0,44	0,45	0,48	
	800	0,36	0,41	0,36	0,34	0,38	0,32	0,31	0,38	0,44	0,34	
15 marcas	50	4,03	3,36	2,87	2,90	3,53	3,05	2,58	4,73	3,11	3,07	
	100	1,29	1,32	1,44	1,30	2,94	1,53	1,49	2,75	2,63	2,70	
	150	0,87	0,95	0,85	0,78	0,91	0,92	0,98	0,89	0,98	0,93	
	200	0,61	0,83	0,81	0,65	0,60	0,90	0,78	0,72	0,64	0,76	
	400	0,41	0,36	0,35	0,41	0,35	0,39	0,35	0,38	0,35	0,38	
	600	0,26	0,22	0,22	0,26	0,22	0,24	0,20	0,21	0,25	0,20	
	800	0,19	0,15	0,19	0,18	0,20	0,16	0,20	0,18	0,19	0,17	
20 marcas	50	1,80	2,26	9,42	3,94	2,01	3,61	3,51	2,86	5,89	6,23	
	100	0,95	0,89	0,98	0,87	0,91	0,89	0,80	0,94	0,88	0,98	
	150	0,63	0,50	0,65	0,60	0,57	0,72	0,50	0,61	0,55	0,65	
	200	0,45	0,50	0,38	0,49	0,53	0,41	0,51	0,37	0,48	0,39	
	400	0,24	0,24	0,23	0,23	0,22	0,22	0,23	0,23	0,24	0,21	
	600	0,15	0,15	0,12	0,16	0,12	0,15	0,12	0,14	0,15	0,15	
	800	0,12	0,11	0,10	0,10	0,10	0,11	0,12	0,11	0,10	0,11	

[†] indica que a primeira linha está vazia por falta de informações necessárias para a geração de grupos de ligação com 50 indivíduos e concentração de cinco marcas por grupo de ligação.

Como comentado, em relação à confiabilidade, quanto menores os valores de variância dos valores entre marcas adjacentes e de desvio-padrão entre repetições, maior será a confiabilidade dos dados.

A redução nos valores de variância com o aumento no tamanho da população é evidenciada pelas diferenças significativas dadas pelo teste de

médias, em que as menores médias estão associadas às maiores populações em todos os grupos de ligação, dentro do nível de saturação de 10 marcas por grupo.

Em relação ao nível de saturação de 15 marcas por grupo de ligação, os valores de variância média das distâncias entre marcas adjacentes também diminuíram com o aumento no tamanho da população, como aconteceu com as populações nos níveis de saturação com 5 e 10 marcas por grupo de ligação. Como exemplo, nesse nível de saturação pode ser observada a diminuição nos valores de variância no grupo de ligação 10, passando de 13,74 na população com 50 indivíduos para 0,84 naquela com 800 indivíduos (Quadro 22). Em relação ao desvio-padrão, apresentam-se 3,07 e 0,17, respectivamente, nas populações com 50 e 800 indivíduos (Quadro 23).

Assim como nos níveis de saturação comentados anteriormente, no nível de saturação com 20 marcas por grupo de ligação os valores de variância diminuíram com o aumento no tamanho das populações, como indicado pelo teste de médias (Quadro 22). Os valores de desvio-padrão (Quadro 23) também indicaram diminuição com o aumento no número de indivíduos na população.

As médias gerais apresentaram diferença significativa nos tamanhos de população, evidenciando-se que as menores populações exibem maiores valores de variância das distâncias entre marcas adjacentes, portanto levando à menor precisão do mapeamento genético.

Em resumo, o que pode ser observado em relação às médias das variâncias é que, à medida que o tamanho da população aumenta, o valor da variância diminui. Além disso, os valores de variância são afetados pelo nível de saturação utilizado, ou seja, reduzindo o nível de saturação, os valores das médias das variâncias aumentam. Num mesmo tamanho de população, aumentando apenas o nível de saturação, pode-se observar que os valores de variância diminuem, variando de 11,47 na população com 100 indivíduos, no nível de saturação com cinco marcas por grupo de ligação, até 5,32 no nível de saturação com 20 marcas por grupo de ligação.

4.5.6. Estresse

Assim como feito nas populações F_2 , o estresse foi utilizado por expressar o grau de concordância dos valores de distância entre cada par de marcas adjacentes em grupos de ligação obtidos no mapeamento. Nesse caso, isso foi feito nas populações de retrocruzamento simuladas com as distâncias dos respectivos pares de marcas no genoma utilizado para simulação das populações.

Vale reforçar que os valores de estresse apresentam significados diferentes em níveis de saturação diferentes, ou seja, devem ser comparados níveis de estresse entre genomas com o mesmo nível de saturação, uma vez que valores iguais de estresse em níveis de saturação diferentes têm significados diferentes, não sendo recomendável a comparação de estresses iguais de níveis de saturação diferentes.

Para obtenção dos valores médios de estresse, foi feita uma média aritmética dos valores de estresse obtidos nas repetições em que houve a formação de 10 grupos de ligação no mapeamento. Os valores obtidos são apresentados no Quadro 24, juntamente com o resultado de comparações feitas entre médias.

Analisando-se os efeitos do tamanho da população dentro do nível de saturação de genoma com cinco marcas por grupo de ligação, ainda não é observável a tendência à redução nos valores de estresse médio nos grupos de ligação com o aumento no tamanho da população segregante, como observado nas populações F_2 nesse mesmo nível de saturação. Essa tendência pode ser vista apenas nas médias gerais dos valores médios de estresse representados na última coluna do Quadro 24, em que o valor do estresse passa de 11,46 para 5,79%, considerando-se as populações com 150 e 800 indivíduos, respectivamente. Os valores médios de estresse das populações com 50 indivíduos não foram apresentados por não ter nenhuma repetição com 10 grupos de ligação. Quanto às populações com 100 indivíduos, apenas uma repetição apresentou valores de estresse em todos os grupos de ligação, e esses valores são apresentados no Quadro 24, apenas para se ter uma idéia dos valores que poderiam ter sido encontrados, mas não servem para análises mais profundas.

Estimativas de valores médios de estresse de 11,46 e 5,79 implicam, em termos de desvio médio, valores de 1,38 e 0,62, respectivamente (Quadro 25). Além da redução nos valores de estresse médio, ocorreu ligeira redução na amplitude de variação dos valores de estresse entre repetições, à medida que aumentou o tamanho das populações segregantes. Nesse nível de saturação, não se pode concluir sobre a confiabilidade e permanência dessa tendência, em razão do baixo nível de saturação e das poucas repetições analisadas.

Quadro 25 – Desvio-padrão dos valores de estresse médios nas repetições que recuperaram 10 grupos de ligação, em populações F_2 , avaliados em quatro níveis de saturação

Saturação	T.P.	Grupos de Ligação										
		1	2	3	4	5	6	7	8	9	10	
5 marcas	50	-	-	-	-	-	-	-	-	-	-	¹
	100	0	0	0	0	0	0	0	0	0	0	
	150	4,39	4,36	3,72	3,67	3,24	3,49	5,95	3,90	2,73	5,23	
	200	5,37	3,45	3,05	3,50	3,98	3,98	4,15	3,92	4,87	3,60	
	400	2,84	2,45	2,73	2,68	2,62	2,76	2,95	2,86	2,85	2,59	
	600	2,55	2,31	2,56	2,32	2,19	2,22	2,44	2,24	2,50	2,45	
	800	2,40	2,18	1,85	1,75	2,18	1,98	2,15	1,95	1,93	2,08	
10 marcas	50	7,69	9,36	9,83	10,27	9,03	8,04	7,47	8,32	9,18	7,69	
	100	7,38	6,92	8,02	6,42	5,98	6,40	7,84	6,58	6,40	6,08	
	150	5,72	6,03	5,14	5,56	5,14	5,80	5,24	5,67	6,32	5,83	
	200	5,34	5,53	4,87	4,79	4,58	4,33	4,63	4,65	4,58	4,85	
	400	3,54	3,44	3,42	3,38	3,46	3,20	3,66	3,24	3,45	3,44	
	600	2,73	2,80	2,82	2,72	2,74	2,77	2,68	3,31	2,88	2,86	
	800	2,45	2,22	2,68	2,47	2,56	2,47	2,47	2,46	2,48	2,80	
15 marcas	50	10,34	10,93	13,28	7,78	11,79	10,36	9,62	10,87	11,10	9,76	
	100	6,36	7,75	7,07	7,92	6,47	6,06	6,60	6,78	6,29	6,64	
	150	6,35	6,03	5,38	5,11	5,35	5,31	5,65	5,18	6,35	6,44	
	200	5,53	5,01	4,77	4,98	5,00	5,05	5,92	5,09	4,95	5,19	
	400	4,14	3,59	3,49	3,83	3,56	3,91	3,53	3,22	3,50	3,71	
	600	3,05	2,81	2,84	3,01	2,54	2,81	2,76	2,96	2,59	2,87	
	800	2,50	2,79	2,58	2,33	2,63	2,79	2,48	2,49	2,57	2,43	
20 marcas	50	12,66	9,91	11,88	10,7	11,70	11,71	12,75	12,69	12,17	10,97	
	100	6,69	6,96	6,72	7,49	7,78	7,40	7,05	7,77	7,05	7,27	
	150	5,79	5,96	5,66	5,83	5,33	6,07	6,34	5,98	7,04	4,95	
	200	6,08	4,66	5,77	5,17	5,15	4,91	4,58	5,53	5,03	5,23	
	400	3,77	4,01	3,33	3,32	3,59	3,06	3,39	3,35	3,57	3,67	
	600	2,90	2,86	2,64	2,92	2,78	3,43	2,99	2,97	2,94	2,77	
	800	2,70	2,43	2,47	2,50	2,58	2,82	2,37	2,51	2,38	2,52	

¹ indica que a primeira linha está vazia por falta de informações necessárias para a geração de grupos de ligação com 50 indivíduos e concentração de cinco marcas por grupo de ligação.

Analisando os efeitos do tamanho da população dentro do nível de saturação de genoma com 10 marcas por grupo de ligação e tomando como exemplo o grupo de ligação 3, cujos valores médios de estresse variam de 36,64 (a) a 10,19% (d) nas populações com 50 e 800 indivíduos, respectivamente, já é possível notar uma tendência à diminuição nos valores médios de estresse com o aumento no tamanho das populações, e esse comportamento é seguido pelos demais grupos de ligação, como podem ser observados no Quadro 24. Os valores médios de estresse das populações com 50 indivíduos, à semelhança do ocorrido nas populações F_2 com o mesmo nível de saturação, diferem de todas as outras com maior número de indivíduos. Em populações com 100 indivíduos, essa diferença significativa ocorre na maioria dos grupos de ligação, a não ser nos grupos 5 e 10. Quanto aos valores médios de estresse dos demais tamanhos de populações, podem ser separados em grupos, dentro dos quais os valores não diferem significativamente entre si, pelo teste de médias. Um exemplo é o grupo de ligação 3, em que as médias formaram um grupo com a população com 50 indivíduos (a), outro com a população com 100 indivíduos (b), outro com as populações com 150 e 200 indivíduos (c) e um último com as populações com 400, 600 e 800 indivíduos (d). Em relação à média geral dos valores médios de estresse, existe diferença significativa entre as médias de todos os tamanhos de população, como se pode observar na última coluna do Quadro 24.

Quanto ao nível de saturação com 15 marcas por grupo de ligação, à semelhança do que aconteceu no caso anterior, os valores médios de estresse das populações com 50 e 100 indivíduos, em todos os grupos de ligação, diferiram entre si e das médias das demais populações. Os valores médios de estresse dos demais tamanhos de população em todos os grupos de ligação, como no nível de saturação anterior, podem ser classificados em dois ou três grupos, dentro dos quais os valores médios de estresse não diferiram significativamente. Os valores das médias gerais dos valores médios de estresse, apresentados na última coluna do Quadro 24, apresentaram diferença significativa em todos os valores. Essa diferença é uma mostra mais forte do que já pode ser percebida na análise dos grupos de ligação. Existe tendência à diminuição dos valores de estresse com o aumento no tamanho das populações segregantes analisadas. Tomando como

exemplo os valores das médias gerais, têm-se 49,84% (a) na população com 50 indivíduos e 13,09% (g) na de 800 indivíduos.

Quanto aos valores de desvio-padrão, relacionados aos valores médios de estresse do nível de saturação com 15 marcas por grupo de ligação, ocorreu pequena redução quando aumentou o tamanho da população analisada; isso só não aconteceu em dois tamanhos de população, nos grupos de ligação 4, 5, 7 e 9.

No nível de saturação com 20 marcas por grupo de ligação, é evidente a diminuição nos valores médios de estresse em todos os grupos de ligação, à medida que aumenta o tamanho das populações. A diminuição só não foi constatada nas populações com 600 e 800 indivíduos dos grupos de ligação 2, 5 e 7, em que os valores médios de estresse não diferiram significativamente entre si, pelo teste de médias. Quanto às médias gerais, última coluna do Quadro 24, todas diferiram significativamente entre si, pelo teste de Tukey a 1%. Nesse nível, a variação foi de 59,02 (a) a 15,41% (g), nas populações com 50 e 800 indivíduos, respectivamente.

No geral, a redução nos valores de estresse com o aumento no tamanho das populações dentro de cada nível de saturação é evidenciada pelas diferenças significativas entre os valores médios obtidos com a utilização do teste de comparação das médias, pelo teste de Tukey a 1%, em todos os grupos de ligação. Nas populações de retrocruzamento, assim como nas populações F_2 , os menores valores de estresse estão associados às maiores populações e os maiores valores de estresse, às menores populações. Mesmo no menor nível de saturação, embora de maneira não muito clara, essa associação pode ser percebida.

A redução na amplitude de variação também pode ser vista pela análise dos valores de desvio-padrão (Quadro 25), que, assim como os valores médios de estresse, no geral diminuem com o aumento no tamanho das populações analisadas. Essa diminuição evidencia maior confiabilidade no mapeamento de populações com maior número de indivíduos. Como último exemplo e considerando o grupo de ligação 3 e o maior nível de saturação, tem-se que o desvio-padrão foi de 11,68 e 2,47, respectivamente nos tamanhos de população de 50 e 800 indivíduos.

Segundo Soller e Beckmann (1983), quando marcadores co-dominantes estão disponíveis, análises baseadas em gerações F_2 serão mais úteis que aquelas com base em gerações de retrocruzamento, por fornecerem informações tanto em relação à dominância quanto ao efeito maior do QTL identificado. Isso também pode ser uma explicação para a necessidade de maior tamanho de população para a obtenção de mapas confiáveis em populações de retrocruzamento, ressaltando-se que os dados foram simulados considerando apenas marcas co-dominantes.

5. CONCLUSÕES

Para obtenção de informações suficientes, de modo que sejam gerados mapas confiáveis, análises devem estar associadas à utilização de tamanho de amostra e número de marcas adequados. Mapas com sérias distorções serão obtidos com utilizações de uma quantidade inadequada de indivíduos, mesmo com grande quantidade de marcas e vice-versa, pois, quando o número de indivíduos analisados é muito pequeno, só o aumento no número de marcas não é suficiente para evitar anomalias. Mesmo que seja usado grande número de indivíduos, se a quantidade de marcas for muito pequena também serão gerados mapas com distorções. Em relação a isso, pode ser dito que é fortemente recomendável a não-utilização de populações com 50 indivíduos ou menos para geração de mapas, tanto em populações F_2 quanto em populações de retrocruzamento, mesmo com uma saturação de 20 marcas por grupo de ligação.

A maior confiabilidade dos mapas não aumenta indefinidamente com o aumento no tamanho da população e na quantidade de marcas utilizadas, sendo que a partir de determinado tamanho, 400 indivíduos nesse caso, aumento no número de marcas tem menor influência na melhoria da qualidade dos mapas. Do mesmo modo, quando já se trabalha com grande quantidade de marcas, o aumento no tamanho da população também tem sua influência reduzida.

A escolha do número mínimo de indivíduos a ser utilizado para a recuperação do número exato de grupos de ligação do genoma deve estar associada ao número de marcadores disponíveis. Em populações F_2 , tamanhos próximos de 100, 150 e 200 indivíduos podem ser utilizados para uma recuperação completa do genoma, no caso de saturação de 5; 6,66; e 10 cM, respectivamente. No caso de populações de retrocruzamento, as amostras devem ter o tamanho mínimo de 200 indivíduos nos três primeiros níveis de saturação e de 800 indivíduos no menor nível de saturação.

No caso de existirem quantidade limitada de indivíduos, por volta de 100, para que seja feita uma avaliação mapas confiáveis podem ser gerados, desde que se utilize uma saturação superior a 5 cM. No caso de populações de retrocruzamento, nesse mesmo nível de saturação deveriam ser usados pelo menos 150 indivíduos.

REFERÊNCIAS

- BECKMANN, J. S.; SOLLER, M. Restriction fragment length polymorphisms in genetic improvement: methodologies. Mapping and costs. **Theor. Appl. Genet.**, v. 65, p. 35-43, 1983.
- BORÉM, A. **Melhoramento de plantas**. Viçosa, MG: Editora UFV, 1997. 547 p.
- BRAHM, L.; RÖCHER, T.; FRIEDT, W. PCR-based markers facilitating marker assisted selection in sunflower for resistance to downy mildew. **Crop Sci.**, v. 40, p. 676-682, 2000.
- BRUNELLI, K.R.; SILVA, H.P.; CAMARGO, L.E.A. Mapeamento de genes de resistência a *Puccinia polysora* em milho. **Fitopatologia Brasileira**, v. 27, p. 134-140, 2002.
- CARBONELL, E.A.; ASINS, M.J.; BASELGA, M.; BALANSARD, E.; GERIG, T.M. Power studies in the estimation of genetic parameters and the localization of quantitative trait loci for backcross and doubled haploid populations. **Theor. Appl. Genet.**, v. 86, p. 411-416, 1993.
- CLARKE, G.M. **Statistics and experimental design**: an introduction for biologists and biochemists. 3. ed. New York-Toronto, 1994. 280 p.
- CONTI, J.H.; MINAMI, K.; GOMES, L.H.; TAVARES, F.C.A. Estimativa da similaridade genética e identificação de cultivares de morangueiro por análise de RAPD. **Horticultura Brasileira**, Brasília, v. 20, n. 2, p. 145-152, junho 2002.

CORRÊA, F.J. de C. **Avaliação de métodos de seleção tradicionais, assistida por marcadores moleculares e por genes candidatos, com dados simulados.** Viçosa, MG: UFV, 2001. 54 f. Dissertação (Mestrado em Zootecnia) – Universidade Federal de Viçosa, Viçosa.

CORRÊA, R.X.; ABDELNOOR, R.V.; FALEIRO, F.G.; CRUZ, C.D.; MOREIRA, A.M.; BARROS, E.G. de. Genetic distances in soybean base don RAPD markers. **Bragantia**, Campinas, v. 58, n. 1, p. 15-22, 1999.

CORRÊA, R.X.; GOOD-GOD, P.I.V.; OLIVEIRA, M.L.P.; NIETSCH, S.; MOREIRA, M.A.; BARROS, E.G. Herança da resistência à mancha-angular do feijoeiro e identificação de marcadores moleculares flanqueando o loco de resistência. **Fitopatologia Brasileira**, v. 26, p. 27-32, 2000.

CRUZ, C.D. **Programa para análise de dados moleculares e quantitativos – GQMOL.** Viçosa, UFV, 2005. (Software).

CRUZ, C.D. **GENES:** programa de análise e processamento de dados baseados em modelos de genética e estatística experimental. Versão 2.1. Viçosa, MG, UFV, 2004.

CRUZ, C.D.; CARNEIRO, P.C.S. **Modelos biométricos aplicados ao melhoramento genético.** Viçosa, MG: Editora UFV, 2003. 585 p. v. 2.

DACHS, J.N.W. **Estatística computacional.** Rio de Janeiro: LTC Editora, 1988. 236 p.

DALBÓ, M.A.; YE, G.N.; WEEDEN, N.F.; WILCOX, W.F.; REISCH, B.I. Marker-assisted selection for powdery mildew resistance in grapes. Department of Horticultural Science, New York State Agricultural Experiment Station. **Geneva**, Cornell University, 14456, v. 126, n. 1, p. 83-89, 2001.

DARVASI, A.; WEINREB, A.; MINKE, V.; WELLER, J.I.; SOLLER, M. Detecting marker-QTL linkage and estimating QTL gene effect and map location using a saturated genetic map. **Genetics**, v. 134, p. 943-951, July, 1993.

DEMPSTER, A.B.; LAIRD, N.M.; RUBIN, D.B. Maximum likelihood from incomplete data via the EM algorithm. **Journal Royal Statistical Society (B)**, v. 39, p. 71361-71398, 1977.

FALCÃO, C.L.; PAPPAS, M.C.R.; LOURENÇO, R.T.; ALENCAR, M.M.; BATISTA, A.R.S.; PAPPAS Jr., G.J.; GRATTAPAGLIA, D. **Desenvolvimento e mapeamento de microssatélites derivados de ESTs em *Eucalyptus*.** Brasília, DF: EMBRAPA, dezembro, 2004. (Circular Técnica, 32).

FALCONER, D.S. **Introdução à genética quantitativa**. Tradução de M.A. Silva e J.C. SILVA. Viçosa, MG: UFV, Impr. Univ., 1987. 279 p.

FALEIRO, F.G.; RAGAGNIN, V.A.; SCHUSTER, I.; CORRÊA, R.X.; GOOD-GOD, P.I.; BROMMONSHENKEL, S.H.; MOREIRA, M.A.; BARROS, E.G. Mapeamento de genes de resistência do feijoeiro à ferrugem, antracnose e mancha-angular usando marcadores RAPD. **Fitopatologia Brasileira**, v. 28, p. 059-066, 2003.

FERREIRA, M.E.; GRATAPAGLIA, D. Classes de marcadores moleculares para análise genética. In: FERREIRA, M.E.; GRATAPAGLIA, D. **Introdução ao uso de marcadores em análise genética**. 2. ed. Brasília: MA/EMBRAPA/CENARGEN, 1996. p. 13-67.

FERREIRA, M.E.; GRATTAPAGLIA, D. **Introdução ao uso de marcadores moleculares em análise genética**. Brasília: EMBRAPA CENARGEN, 1998. 220 p.

FRISH, M.; BOHN, M.; MELCHINGER, A. E. Comparison of selection strategies for marker-assisted backcrossing of a gene. **Crop Sci.**, v. 39, p. 1295-1301, 1999.

GANGA, R.M.D.; RUGGIERO, C.; LEMOS, E.G. de M.; GRILI, G.V.G.; GONÇALVES, M.M.; CHAGAS, E.A.; WICKERT, E. Diversidade genética em maracujazeiro-amarelo utilizando marcadores moleculares fAFLP. **Rev. Bras. Frutic.**, Jaboticabal, v. 26, n. 3, p. 494-498, dezembro, 2004.

GRATTAPAGLIA, D.; SEDEROFF, R. Genetic linkage maps of *Eucalyptus grandis* and *E. urophylla* using a pseudo-testcross mapping strategy and RAPD markers. **Genetics**, v. 137, p. 1121-1137, 1994.

HALDANE, J. B. S. The combination of linkage values and the calculation of distances between the loci of linked factors. **J. Genet.**, v. 8, p. 299-309, 1919.

HOSPITAL, F.; CHEVALET, C.; MULSANT, P. Using markers in gene introgression breeding programs. **Genetics**, v. 132, p. 1199-1210, december, 1992.

KNOTT, S. A.; HALEY, C. S. Aspects of maximum likelihood methods for the mapping of quantitative trait loci in line crosses. **Genet. Res.**, Camb., v. 60, p. 139-151, 1992.

KONDO, T.; TERADA, K.; HAYASHI, E.; KURAMOTO, N.; OKAMURA, M.; KAWASAKI, H. **RAPD markers linked to a gene for resistance to pine needle gall midge in Japanese black pine (*Pinus thunbergii*)**. [S.l.]: Springer-Verlag, 2000. p. 391-395.

KOSAMBI, D. D. The estimation of map distances from recombination values. **Ann Eugen.**, v. 12, p. 172-175, 1944.

LANDER, E. S.; GREEN, P.; ABRAHAMSON, J.; BARLOW, A.; DALY, M. J.; LINCOLN, S. E.; NEWBURG, L. MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. **Genomics**, v. 1, n. 2, p. 174-81, Oct. 1987.

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

LANNES, S. D.; ZIMMER, P. D.; OLIVEIRA, A. C. de; IRAJÁ, F.; CARVALHO, F. de; VIEIRA, E. ^a; JUNIOR, A. M. de M.; KOPP, M. M.; FREITAS, F. A. de. Regeneração *in vitro* de anteras de arroz irrigado (*Oryza sativa* L.) e mapeamento de QTL associado. **Ciência Rural**, v. 34, n. 5, p. 1355-1362, set.-out. 2004.

LEBRETON, C.M.; VISSCHER, P.M. Empirical nonparametric bootstrap strategies in quantitative trait loci mapping: Conditioning on the genetic model. **Genetics Society of America**, v. 148, p. 525-535, Jan. 1998.

LIU, B.H. **Statistical genomics, linkage, mapping and QTL analysis**. Boca Raton: CRC Press, 1998. 611 p.

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

MARTÍNEZ, O.; CURNOW, R. N. Estimating the locations and the size of the effects of quantitative trait using flanking markers. **Theor. Appl. Genet.**, v. 85, p. 480-488, 1992.

MENDEL, G. Experiments in plant hybridization. Tradução de William Bateson 1901.[on line]. Disponível em: <<http://www.esp.org/foundations/genetics/classical/gm-64.pdf>>. Acesso em: 15 Ago. 2002.

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

MORGAN, T. H. Sex limited inheritance in *Drosophila*. **Science**, v. 32, p. 120-122, 1910.

NASS, L. L.; VALOIS, A. C. C.; MELO, I. S. de; VALADARES-INGLIS, M. C. **Recursos genéticos & melhoramento**. Rondonópolis, MT: Fundação MT, 2001. 1183 p.

OOIJEN, J. W.V. Accuracy of mapping quantitative trait loci in autogamous species. **Theor. Appl. Genet.**, v. 84, p. 803-8011, 1992.

PATERSON, A. H.; LANDER, E.; HEWITT, J. D.; PETERSON, S.; LINCLIN, S. E.; TANKSLEY, S. D. Resolution of quantitative traits into mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. **Nature**, v. 335, p. 721-726, 1988.

ROCHA, R. B.; PEREIRA, J. F.; CRUZ, C. D.; QUEIROZ, M. V. de; ARAÚJO, E. F. de. O mapeamento genético no melhoramento de plantas. **Revista Biotecnologia Ciência e Desenvolvimento**, v. 30, p. 27-31, 2003.

SAX, K. Association of size differences with seed-coat pattern and pigmentation in *Phaseolus vulgaris*. **Genetics**, v. 8, p. 552-560, 1923.

SCHUSTER, I.; ABDELNOOR, R. V.; MARIN, S. R. R.; CARVALHO, V.P.; KIILH, R. A. S.; SILVA, J. F. V.; SEDIYAMA, C. S.; BARROS, E. G.; MOREIRA, M. A. Identification of a new major QTL associated with resistance to soybean cyst nematode (*Heterodera glycines*). **Theor. Appl. Genet.**, v. 102, p. 91-96, 2001.

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

SILVA, A.R. da. **Análise genética de caracteres quantitativos em milho com delineamento III e marcadores moleculares**. Piracicaba, SP: ESALQ, 2002. 143 f. (Doutorado) – Escola Superior de Agricultura Luiz de Queiroz, Piracicaba.

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

SOLLER, M.; BECKMANN, J.S. Genetic polymorphism in varietal identification and genetic improvement. **Theor. Appl. Genet.**, v. 67, p. 25-33, 1983.

STUTERVANT, A. H. **A history of genetics**. New York: Harper& Row, 1965. 156 p.

STUTERVANT, A. H. The linear arrangement of six-linked factors in *Drosophila*, as shown by their mode of association. **Journal of Experimental Zoology**, v. 14, p. 43-59, 1913.

TANKSLEY, D.S.; YOUNG, N.D.; PATERSON, A.H.; BONIERBALE, M.W. Reviw. RFLP mapping in plant breeding: new tools for an old science. **Bio/Technology**, v. 7, n. 1, p. 257-264, 1989.

TELLES, M.P. de C.; MONTEIRO, M.S.R.; RODRIGUES, F.M.; SOARES, T.N.; RESENDE, L.V.; AMARAL, A. das G.; MARRA, P.R. Marcadores RAPD na análise da divergência genética entre raças de bovinos e número de locos necessários para a estabilidade da divergência estimada. **Ciência Animal Brasileira**, v. 2, n. 2, p. 87-95, jul./dez. 2001.

VIDOR, M.A.; RUIZ, C.P.; MORENO, S.V.; FLOSS, P.A. Marcadores moleculares em estudos de caracterização de erva-mate (*Ilex paraguariensis* St. Hil.): o sabor. **Cienc. Rural**, Santa Maria, v. 32, p. 3, may/jun. 2002.

VISSCHER, P. M.; HALEY, C. S.; THOMPSON, R. Marker-assisted introgression in backcross breeding programs. **Genetics**, v. 144, p. 1923-1932, december, 1996.

WELLER, J. I. Maximum-likelihood techniques for the mapping and analysis of quantitative trait loci with the aid of genetic markers. **Biometrics**, v. 42, p. 627-640, 1986.

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