

Tamanho de população ideal para mapeamento genético em famílias de irmãos completos

Leonardo Lopes Bhering⁽¹⁾ e Cosme Damião Cruz⁽¹⁾

⁽¹⁾Universidade Federal de Viçosa, Departamento de Biologia Geral, CEP 36571-000 Viçosa, MG. E-mail: leobhering@yahoo.com.br, cdcruz@ufv.br

Resumo – O objetivo deste trabalho foi avaliar o tamanho ótimo de populações de irmãos completos, para estudos de mapas de marcadores moleculares, por meio de simulações de genomas e tamanhos de populações. Foram simulados genomas parentais e amostras de populações de família de irmãos completos do tipo completamente informativas, e também não completamente informativas. As amostras geradas foram de 100, 200, 400 e 600 indivíduos, com três grupos de ligação cada, e 11 marcas moleculares codominantes e multialélicas, espaçadas a dez centimorgans por grupo de ligação. Foram realizadas 100 repetições por amostra. Para populações completamente informativas, o tamanho populacional de 200 indivíduos é suficiente para recuperar as informações originais, contudo, para a população não completamente informativa, é necessária a utilização de uma população maior, de 600 indivíduos.

Termos para indexação: amostragem, genômica, mapeamento, populações exogâmicas.

Optimum population size for genetic mapping in full sibling families

Abstract – The objective of this work was to evaluate the optimum size of populations for the study of genetic mapping of full sibling families, through data simulation of genome and populations. Parental genomes and population samples of full sibling families of both completely and non-completely informative types, were simulated. The generated samples had 100, 200, 400 and 600 individuals, with three linkage groups each, and 11 codominant multi-allelic molecular marks, spaced by ten centimorgans in each linkage group. One-hundred repetitions were accomplished by sample. In completely informative populations, the optimum size of 200 individuals is enough to rescue the original information, however, for the non-completely informative population, it is necessary a larger population, with 600 individuals.

Index terms: sampling, genomics, exogamic populations, mapping.

Introdução

O mapeamento genético facilita o trabalho de melhoramento, uma vez que uma ou mais marcas do genótipo podem estar associadas a um ou mais genes controladores de características qualitativas e quantitativas (QTL). Desse modo, tendo-se o genótipo mapeado, o trabalho de melhoramento pode ser otimizado, tanto na eficiência do programa, quanto na velocidade de obtenção de ganhos, pois é possível a realização de seleção com base nos marcadores. Entretanto, a disponibilidade de um mapa genético fidedigno depende de vários fatores como o tipo de marcador utilizado, o tipo de população analisada e o tamanho da população, além de outros.

Um dos fatores de fundamental importância, para se obter dados consistentes em um trabalho de mapeamento, é o tamanho da amostra ou população a ser estudada. A resolução do mapa e a capacidade de se determinar a seqüência de marcadores nele estão diretamente relacionadas ao tamanho da amostra ou população. Segundo Young (1994), amostras com menos de 50 indivíduos, provavelmente, terão baixa resolução de mapeamento, principalmente na detecção de QTL de pequeno efeito. É fundamental que se estimem adequadamente distâncias entre os genes, estabelecendo-se o ordenamento correto e formando-se grupos de ligações que reflitam o número básico de cromossomos da espécie. Como os eventos de permutação ocorrem ao acaso ao longo do cromossomo, a probabilidade de

recombinação é maior para locos que se encontram a uma maior distância entre si do que para aqueles 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 o cálculo de distância e ordenamento dos genes (ou marcadores) nos cromossomos (Schuster & Cruz, 2004).

Confirmada a existência de ligação entre duas marcas, é indispensável adotar métodos quantitativos para estudar o grau de associação. O método de máxima verossimilhança é utilizado no mapeamento genético para a obtenção de várias estimativas, inclusive as da frequência de recombinação (Liu, 1998).

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

O estabelecimento de mapas genéticos em populações exogâmicas apresenta determinadas complicações que não são encontradas quando utilizados delineamentos a partir de linhagens endogâmicas, como as populações F_2 , de duplo-haplóides, e das oriundas de retrocruzamentos, entre outras. Em populações derivadas do cruzamento entre linhagens endogâmicas, todos os locos segregarão para apenas dois alelos. Além disso, a fase de ligação do duplo heterozigoto pode ser claramente determinada, com base na análise da segregação dos gametas recombinantes da população (Lynch & Walsh, 1998).

Em determinadas espécies de plantas, não é possível obter populações segregantes derivadas de linhagens endogâmicas, em virtude da auto-incompatibilidade, depressão endogâmica, ou do longo período juvenil. Assim, em tais espécies, é preciso utilizar delineamentos experimentais de populações exogâmicas, como famílias de meios-irmãos e de irmãos completos.

Diferentes configurações de marcadores podem estar em segregação, em famílias de irmãos completos originadas a partir do cruzamento entre parentais derivados de população exogâmica. De acordo com Lynch & Walsh (1998), existem três tipos de acasalamentos quanto ao grau de informação da progênie: famílias derivadas de cruzamentos completamente informativos; famílias derivadas de retrocruzamentos; e famílias derivadas de intercruzamentos. Considerando-se dois locos em segregação e marcas codominantes, 81 configurações podem surgir da combinação dos diferentes tipos de acasalamentos citados anteriormente. Destas

configurações, 17 devem proporcionar informação sobre ligação, e o restante não deve conter informação de ligação. Para marcadores dominantes, sete de nove configurações podem proporcionar informação de ligação (Liu, 1998).

O objetivo deste trabalho foi avaliar o tamanho ótimo de populações de irmãos completos, para estudos de mapas de marcadores moleculares, por meio de simulações de genomas e tamanhos de populações.

Material e Métodos

Foram simulados genomas parentais e amostras de populações de família de irmãos completos do tipo completamente informativa, e também de famílias de irmãos completos não completamente informativas, pelo módulo de simulação do programa para análise de dados moleculares e quantitativos – GQMOL (Cruz, 2005). As amostras geradas foram de tamanho 100, 200, 400 e 600 indivíduos, com três grupos de ligação cada. Foi gerado o genoma com saturação de 11 marcas moleculares, espaçadas de 10 cM, codominantes, portanto, cada grupo de ligação teve o tamanho de 100 cM. Foram realizadas 100 repetições por amostra, num total de 800 simulações, 400 para famílias completamente informativas e 400 para não completamente informativas.

Na simulação dos genitores para marcadores completamente informativos, foram considerados genitores do tipo $A_iA_j \times A_kA_l$, em que, i, j, k, l são diferentes alelos presentes na população, para originar as famílias de irmãos completos (FIC). Na simulação dos genitores não completamente informativos, para formar a segunda população de irmãos completos, considerou-se a presença de quatro alelos na população genitora tomados ao acaso, com frequências iguais de 0,25 para cada alelo.

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

O processo de simulação das famílias de irmãos completos, tanto com locos completamente informativos quanto parcialmente informativos, seguiu os seguintes passos: a partir do genoma simulado, foram construídos os genótipos parentais; para a população completamente informativa, os pais tiveram constituição alélica do tipo 12x34; para a população não completamente informativa, cada pai poderia possuir qualquer uma das dez possíveis combinações alélicas (11, 12, 13, ..., 34, 44);

a partir dos genótipos parentais, foram gerados os gametas para a formação dos indivíduos das populações de irmãos completos. A produção de gametas foi realizada, tendo-se simulado o pareamento dos homólogos e realizado permutas ao longo dos cromossomos, considerando-se a não existência de interferência, nas regiões delimitadas por duas marcas adjacentes. A probabilidade de ocorrência de recombinação numa região entre marcas adjacentes foi dada de acordo com a distância desses marcadores no genoma simulado. Uma maior distância implicou numa maior possibilidade de ocorrência de recombinação.

Após a geração dos dados, foi realizada a análise de segregação de locos individuais. Foram aplicados testes de qui-quadrado, para verificação da razão de segregação em cada marca de todas as populações geradas. Em seguida, foi realizada a estimação da percentagem de recombinação entre pares de marcas, pelo método da máxima verossimilhança.

A descendência da população completamente informativa foi estabelecida a partir do cruzamento de genitor A_1A_2 com o genitor A_3A_4 . A partir desse cruzamento, quatro possibilidades de fase de ligação poderiam acontecer para os dois locos envolvidos: aproximação-aproximação, aproximação-repulsão, repulsão-aproximação, repulsão-repulsão. Dessa forma, foi possível estimar quatro diferentes medidas de distância, tendo-se adotado como a verdadeira fase de ligação entre os locos aquela que proporcionou a maior estimativa de logaritmo decimal (LOD).

A descendência da população não completamente informativa foi estabelecida a partir do cruzamento de genitores em que, para cada loco, era possível a ocorrência de um dos quatro alelos, de forma que haveria para cada genitor a possibilidade de formação de dez tipos de combinações alélicas para formar sua constituição genotípica.

A partir da constituição genotípica de cada genitor, observou-se o tipo de segregação para cada loco, para obtenção do tipo de cruzamento envolvido. É possível a ocorrência teórica de até quatro fases de ligação, e adotou-se como a real fase de ligação entre os marcadores aquela em que foi obtido o maior LOD.

Todos os genomas obtidos foram comparados com o genoma padrão simulado, e constituíram-se de três grupos de ligação e 11 marcas codominantes equidistantes.

As variáveis analisadas foram: número de grupos de ligação obtidos; tamanho dos grupos de ligação; distâncias médias entre marcas adjacentes nos grupos

de ligação; variâncias das distâncias entre marcas adjacentes nos grupos de ligação; estresse (expressa 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); inversão de posição dos marcadores, verificada pela correlação de Spearman. Todas essas comparações foram realizadas no módulo Comparação de genomas, do programa GQMOL (Cruz, 2005).

Nas análises apresentadas, foram utilizadas apenas as repetições em que houve recuperação dos três grupos de ligação no mapeamento genético.

Foram comparados pelo teste de Tukey, a 5% de probabilidade (erro tipo I), com auxílio do programa Genes (Cruz, 2004) as médias das variáveis: tamanho do grupo de ligação, distância média de marcas adjacentes, variância e estresse para cada grupo de ligação obtido para vários tamanhos de população. Também foram comparadas as médias gerais (médias de todos os grupos de ligação), para cada tamanho de população dentro de cada nível de saturação do genoma.

Resultados e Discussão

O número de grupos de ligação esperado no mapeamento das duas populações era três, que era o número de grupos de ligação que se tinha no genoma original usado para a simulação das populações.

Um 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.

A inversão também é fator importante a ser verificado, para se ter idéia da confiabilidade dos dados obtidos. 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.

A população não completamente informativa apresentou muito mais complicações durante o mapeamento, com maior frequência de inversões e maior número de repetições, descartadas deste trabalho por não ter recuperado os três grupos de ligação previamente declarados (Tabela 1). Houve maior acurácia com uma população completamente informativa de 200 indivíduos, do que com uma população não completamente informativa de 600 indivíduos – ambas recuperaram os três grupos de ligação – porém, o percentual de inversões

na população completamente informativa (0,33%) foi inferior ao da população não informativa (2,33%).

O tamanho esperado nos grupos de ligação, após o mapeamento das populações, era de 100 cM, uma vez que esse era o tamanho de cada grupo de ligação no nível de saturação do genoma utilizado para a simulação das populações.

Quanto ao tamanho médio de cada grupo de ligação, obtido nas repetições com três grupos de ligação, verificou-se que não houve diferenças entre as médias dos diferentes tamanhos de população (Tabela 2). Não houve variação significativa no tamanho médio dos

grupos de ligação com o aumento no tamanho da população e com os diferentes tipos de populações. Não se pode, portanto, fazer uma referência de determinado tamanho da população completamente informativa que se aproxime da população não completamente informativa.

Uma das maneiras adicionais de se observar o comportamento dos tamanhos médios dos grupos de ligação se dá pela análise do desvio-padrão. Com o aumento no número de indivíduos em um mesmo nível de saturação de genoma, esperava-se que o desvio-padrão diminuísse. Os valores obtidos para populações completamente informativas foram inferiores àqueles obtidos para populações não completamente informativas, em todos os tamanhos de populações avaliados (Tabela 2). Observou-se, ainda, que o valor obtido para a população completamente informativa de 100 indivíduos aproximou-se muito daquele obtido na população não completamente informativa de 400 indivíduos. Porém, o valor obtido para a população completamente informativa de 100 indivíduos não foi significativamente diferente dos valores da população não completamente informativa de 100, 200 e 400 indivíduos.

A média das distâncias entre marcas adjacentes e seu desvio-padrão dos vários tamanhos de população segregantes, utilizados no mapeamento genético, estão apresentados na Tabela 3. Os valores obtidos do grupo

Tabela 1. Tamanho da população, número de repetições com três grupos de ligação, número de inversões de populações de família de irmãos completos dos tipos completamente informativas e não completamente informativas, e percentagem de inversões observadas.

| Tamanho da população | Repetições | Número de inversões | Percentagem de inversões |
|---|------------|---------------------|--------------------------|
| Populações completamente informativas | | | |
| 100 | 99 | 0 | - |
| 200 | 100 | 1 | 0,33 |
| 400 | 100 | 0 | - |
| 600 | 100 | 0 | - |
| Populações não completamente informativas | | | |
| 100 | 95 | 32 | 11,22 |
| 200 | 96 | 23 | 7,98 |
| 400 | 100 | 23 | 7,66 |
| 600 | 100 | 7 | 2,33 |

Tabela 2. Média e desvio-padrão do tamanho (cM) de três grupos de ligação, em quatro tamanhos de populações de família de irmãos completos dos tipos completamente informativas e não completamente informativas⁽¹⁾.

| Tamanho da população | Grupo de ligação | | | Média geral | Número de dados |
|---|------------------|----------|-------------|-------------|-----------------|
| | 1 | 2 | 3 | | |
| Média aritmética de populações completamente informativas | | | | | |
| 100 | 102,722abc | 102,152a | 104,172a | 103,015a | 99 |
| 200 | 101,891c | 103,244a | 103,458abc | 102,864a | 100 |
| 400 | 102,825abc | 103,025a | 103,829ab | 103,226a | 100 |
| 600 | 102,967abc | 102,586a | 102,788abcd | 102,780a | 100 |
| Média aritmética de populações não completamente informativas | | | | | |
| 100 | 105,081ab | 102,276a | 103,165abc | 103,507a | 95 |
| 200 | 105,192ab | 103,640a | 101,169cd | 103,333a | 96 |
| 400 | 105,455a | 103,546a | 101,548bcd | 103,516a | 100 |
| 600 | 102,099bc | 103,056a | 100,664d | 101,939a | 100 |
| Desvio-padrão de populações completamente informativas | | | | | |
| 100 | 6,895 | 7,193 | 7,356 | 7,148ab | 99 |
| 200 | 5,007 | 6,097 | 5,741 | 5,615bc | 100 |
| 400 | 3,663 | 3,741 | 3,683 | 3,696bc | 100 |
| 600 | 3,026 | 3,060 | 3,229 | 3,105c | 100 |
| Desvio-padrão de populações não completamente informativas | | | | | |
| 100 | 11,007 | 9,574 | 7,943 | 9,508a | 95 |
| 200 | 8,927 | 6,289 | 5,162 | 6,792abc | 96 |
| 400 | 9,831 | 5,870 | 5,233 | 6,978ab | 100 |
| 600 | 5,781 | 4,160 | 3,238 | 4,393bc | 100 |

⁽¹⁾Médias seguidas por letras iguais não diferem entre si pelo teste de Tukey, a 5% de probabilidade.

de ligação 3, da população não completamente informativa, não foram apresentados, uma vez que esse grupo de ligação apresentou apenas dez marcas, pois uma das suas marcas não segregava por ser do tipo $A_iA_i \times A_iA_i$; com isso, esse grupo de ligação apresentou um grau de saturação diferente dos demais grupos de ligação. Então, para evitar problemas na comparação das informações neste trabalho, esses valores foram excluídos.

Como as populações foram simuladas a partir de um genoma pré-determinado, esperava-se que as médias se aproximassem ao máximo da distância do genoma simulado, que foi de 10 cM. Independentemente do tamanho da população, todos os valores ficaram acima dos inicialmente esperados (Tabela 3).

Não foi observada diferença entre as médias em relação aos tamanhos populacionais avaliados, isso indica que a variação no tamanho populacional e os dois tipos de populações avaliados não proporcionaram vantagem significativa na recuperação dos valores do genoma original.

Houve diferença entre os tratamentos avaliados (Tabela 3) com relação aos desvios-padrão das médias das distâncias, entre marcas adjacentes nas repetições que recuperaram três grupos de ligação, avaliados em quatro tamanhos populacionais diferentes nos dois tipos de populações. A população completamente informativa

de 200 indivíduos apresentou comportamento semelhante à população não completamente informativa de 600 indivíduos. Isso dá uma idéia do tamanho necessário de determinado tipo de população para ser mais bem representada. Assim, ao se trabalhar com populações completamente informativas, pode-se usar apenas um terço do tamanho populacional daquele usado para populações não informativas.

A partir das distâncias entre marcas adjacentes obtidas nos grupos de ligação foi estimada a variância amostral (Tabela 4) que é referente aos erros para qualquer tamanho de população dos genomas avaliados, uma vez que, os genomas utilizados para geração das populações segregantes tinham seus marcadores distribuídos de forma equidistante. Quanto menores os valores de variância mais equidistantes estarão distribuídos às 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, indicando uma boa recuperação do genoma com o mapeamento das populações segregantes (Cruz, 2006).

Ao se analisar o efeito do tamanho e dos tipos de população, observou-se que os valores obtidos para a população completamente informativa eram menores do que aqueles obtidos para a população não completamente informativa, tendo-se considerado os mesmos tamanhos

Tabela 3. Média e desvio-padrão das distâncias (cM), entre marcas adjacentes nas repetições que recuperaram três grupos de ligação, em quatro tamanhos de populações de família de irmãos completos dos tipos completamente informativas e não completamente informativas⁽¹⁾.

| Tamanho da população | Grupo de ligação | | | Média geral | Número de dados |
|--|------------------|--------|--------|-------------|-----------------|
| | 1 | 2 | 3 | | |
| Média de populações completamente informativas | | | | | |
| 100 | 10,272 | 10,215 | 10,417 | 10,301a | 99 |
| 200 | 10,189 | 10,324 | 10,345 | 10,286a | 100 |
| 400 | 10,282 | 10,302 | 10,382 | 10,322a | 100 |
| 600 | 10,296 | 10,258 | 10,278 | 10,278a | 100 |
| Média de populações não completamente informativas | | | | | |
| 100 | 10,508 | 10,227 | - | 10,367a | 95 |
| 200 | 10,519 | 10,363 | - | 10,441a | 96 |
| 400 | 10,545 | 10,354 | - | 10,449a | 100 |
| 600 | 10,209 | 10,305 | - | 10,257a | 100 |
| Desvio-padrão de populações completamente informativas | | | | | |
| 100 | 0,689 | 0,719 | 0,735 | 0,714ab | 99 |
| 200 | 0,500 | 0,609 | 0,574 | 0,561bc | 100 |
| 400 | 0,366 | 0,374 | 0,368 | 0,369c | 100 |
| 600 | 0,302 | 0,306 | 0,322 | 0,310c | 100 |
| Desvio-padrão de populações não completamente informativas | | | | | |
| 100 | 1,100 | 0,957 | - | 1,020a | 95 |
| 200 | 0,892 | 0,628 | - | 0,760ab | 96 |
| 400 | 0,983 | 0,587 | - | 0,785ab | 100 |
| 600 | 0,578 | 0,416 | - | 0,497bc | 100 |

⁽¹⁾Médias seguidas por letras iguais não diferem estatisticamente entre si pelo teste de Tukey, a 5% de probabilidade.

populacionais (Tabela 4). Verificou-se, ainda, que o valor para a população completamente informativa de 200 indivíduos foi inferior à estimativa apresentada na população não completamente informativa de 600 indivíduos. Porém, ao se realizar o teste de médias, os valores dessas populações não foram significativamente diferentes. Mais uma vez, percebeu-se a tendência de correlação entre os dados apresentados pela população completamente informativa de 200 indivíduos e a não completamente informativa de 600 indivíduos.

Apesar de os valores obtidos das distâncias entre marcadores pela população completamente informativa serem bem inferiores àqueles obtidos pela população não completamente informativa, a ponto de o maior valor da população completamente informativa (4,324) ser menor do que o menor valor obtido na população não completamente informativa, não houve diferença significativa entre os valores, não tendo sido possível obter maiores conclusões a respeito desses dados (Tabela 4).

Observou-se tendência de redução dos valores de estresse médio com o aumento no tamanho da população, nas duas populações avaliadas (Tabela 5). Os valores

de estresse médio diferiram significativamente, e os valores obtidos na população completamente informativa de 200 indivíduos foram próximos aos valores obtidos na população não completamente informativa de 400 indivíduos, porém, os valores da população completamente informativa de 200 indivíduos não apresentaram diferenças significativas dos valores da população não completamente informativa de 200, 400 e 600 indivíduos, o que tornou difícil observar uma tendência clara de relação entre os tipos de populações e seus tamanhos.

Em relação à amplitude da variação dos valores médios de estresse, observou-se que não houve diferença nos dois tipos de populações no tamanho de 100 indivíduos. A população completamente informativa de 200 indivíduos teve comportamento semelhante às populações não completamente informativas de 200, 400 e 600 indivíduos. A população completamente informativa de 400 indivíduos apresentou comportamento semelhante ao da população não completamente informativa de 600 indivíduos (Tabela 5). Assim, mais uma vez não se pôde concluir uma relação entre os tipos de populações completamente informativas e não completamente informativas. Porém, apesar de não ter

Tabela 4. Médias e desvio-padrão da variância das distâncias entre marcas adjacentes, nas repetições que recuperaram três grupos de ligação, em quatro tamanhos de populações de família de irmãos completos dos tipos completamente informativas e não completamente informativas⁽¹⁾.

| Tamanho da população | Grupo de ligação | | | Média geral | Número de dados |
|--|------------------|--------|-------|-------------|-----------------|
| | 1 | 2 | 3 | | |
| Média de populações completamente informativas | | | | | |
| 100 | 5,327 | 5,707 | 5,462 | 5,498ab | 99 |
| 200 | 2,797 | 3,017 | 2,736 | 2,850b | 100 |
| 400 | 1,245 | 1,236 | 1,259 | 1,246b | 100 |
| 600 | 0,838 | 0,875 | 0,867 | 0,860b | 100 |
| Média de populações não completamente informativas | | | | | |
| 100 | 14,972 | 10,106 | - | 12,539a | 95 |
| 200 | 12,810 | 5,324 | - | 9,067ab | 96 |
| 400 | 11,351 | 3,209 | - | 7,280ab | 100 |
| 600 | 4,953 | 1,928 | - | 3,440ab | 100 |
| Desvio-padrão de populações completamente informativas | | | | | |
| 100 | 4,774 | 4,645 | 3,553 | 4,324a | 99 |
| 200 | 2,111 | 1,899 | 1,636 | 1,882a | 100 |
| 400 | 0,653 | 0,723 | 0,886 | 0,754a | 100 |
| 600 | 0,442 | 0,436 | 0,499 | 0,459a | 100 |
| Desvio-padrão de populações não completamente informativas | | | | | |
| 100 | 15,401 | 5,409 | - | 10,405a | 95 |
| 200 | 16,931 | 2,687 | - | 9,809a | 96 |
| 400 | 17,213 | 2,877 | - | 10,045a | 100 |
| 600 | 10,654 | 1,039 | - | 5,846a | 100 |

⁽¹⁾Médias seguidas por letras iguais não diferem entre si pelo teste de Tukey, a 5% de probabilidade.

Tabela 5. Média e desvio-padrão de estresse (%), nas repetições em que se recuperaram três grupos de ligação, em quatro tamanhos de populações de família de irmãos completos dos tipos completamente informativas e não completamente informativas⁽¹⁾.

| Tamanho da população | Grupo de ligação | | | Média geral | Número de dados |
|--|------------------|--------|--------|-------------|-----------------|
| | 1 | 2 | 3 | | |
| Média de populações completamente informativas | | | | | |
| 100 | 21,942 | 22,852 | 22,712 | 22,502b | 99 |
| 200 | 15,931 | 17,256 | 16,450 | 16,546cd | 100 |
| 400 | 11,187 | 11,152 | 11,431 | 11,257e | 100 |
| 600 | 9,377 | 9,412 | 9,511 | 9,430e | 100 |
| Média de populações não completamente informativas | | | | | |
| 100 | 25,706 | 30,492 | - | 28,099a | 95 |
| 200 | 18,050 | 22,474 | - | 20,262bc | 96 |
| 400 | 15,117 | 17,281 | - | 16,199cd | 100 |
| 600 | 13,386 | 13,734 | - | 13,560de | 100 |
| Desvio-padrão de populações completamente informativas | | | | | |
| 100 | 7,279 | 6,914 | 6,873 | 7,022a | 99 |
| 200 | 4,893 | 4,619 | 4,525 | 4,679bc | 100 |
| 400 | 2,874 | 3,160 | 3,303 | 3,112de | 100 |
| 600 | 2,341 | 2,491 | 2,393 | 2,408e | 100 |
| Desvio-padrão de populações não completamente informativas | | | | | |
| 100 | 7,651 | 8,180 | - | 7,915a | 95 |
| 200 | 6,032 | 5,173 | - | 5,602b | 96 |
| 400 | 5,433 | 6,110 | - | 5,771b | 100 |
| 600 | 4,214 | 3,391 | - | 3,802cd | 100 |

⁽¹⁾Médias seguidas por letras iguais não diferem entre si pelo teste de Tukey, a 5% de probabilidade.

ficado evidente a relação entre esses tipos de populações, ficou claro que em todas as análises se necessitou de tamanho de população completamente informativa menor do que não completamente informativa.

Verificou-se que, para populações completamente informativas, o número mínimo de 200 indivíduos é o mesmo tamanho recomendado por Cruz (2006), ao utilizar populações F_2 e de retrocruzamento, com nível de saturação de 10 cM entre marcadores.

Conclusões

1. Para populações completamente informativas, o tamanho populacional de 200 indivíduos é suficiente para recuperar as informações originais de forma satisfatória, na construção de mapas moleculares.

2. Para populações não completamente informativas, é necessário o tamanho populacional de 600 indivíduos, para se obter um mapa molecular fidedigno.

Agradecimentos

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico, por concessão de bolsa.

Referências

- CRUZ, C.D. **GENES**: programa de análise e processamento de dados baseados em modelos de genética e estatística experimental. Viçosa: UFV, 2001. (Versão 2004. 2.1).
- CRUZ, C.D. **Programa para análise de dados moleculares e quantitativos**: GQMOL. Viçosa: UFV, 2005.
- CRUZ, E.M. **Efeito da saturação e do tamanho de populações F_2 e de retrocruzamento sobre a acurácia do mapeamento genético**. 2006. 119p. Tese (Doutorado) - Universidade Federal de Viçosa, Viçosa.
- LIU, B.H. **Statistical genomics**: linkage, mapping and QTL analysis. Boca Raton: CRC Press, 1998. 611p.
- LYNCH, M.; WALSH, B. **Genetics and analysis of quantitative traits**. Sunderland: Sinauer Associates, 1998. 980p.
- SCHUSTER, I.; CRUZ, C.D. **Estatística genômica** - aplicada a populações derivadas de cruzamentos controlados. Viçosa: UFV, 2004. 568p.
- SILVA, L. da C. **Simulação do tamanho da população e da saturação do genoma para mapeamento genético de RILs**. 2005. 120p. Dissertação (Mestrado) - Universidade Federal de Viçosa, Viçosa.
- YOUNG, N.D. Constructing a plant genetic linkage map with DNA markers. In: PHILLIPS, R.L.; VASIL, I.K. **DNA: based markers in plants**. Dordrecht: Kluwer Academic Publisher, 1994. p.31-47.

Recebido em 20 de agosto de 2007 e aprovado em 3 de março de 2008