

DESEMPENHO FENOTÍPICO AO UTILIZAR DIFERENTES DENSIDADES DE MARCADORES MOLECULARES NO MAPEAMENTO GENÔMICO

PHENOTYPIC PERFORMANCE USING DIFFERENT MOLECULAR MARKERS DENSITY IN GENOMIC MAPPING

Marcelo JANGARELLI¹; Ricardo Frederico EUCLYDES²; Paulo Roberto CECON³

1. Professor, Doutor, Departamento de Matemática, Universidade Federal Rural do Rio de Janeiro, Seropédica, RJ, Brasil. mejanga@hotmail.com; 2. Professor, Doutor, Departamento de Zootecnia, Universidade Federal de Viçosa - UFV, Viçosa, MG, Brasil; 3. Professor, Doutor, Departamento de Estatística - UFV, Viçosa, MG, Brasil.

RESUMO: Objetivou-se avaliar, através do programa de simulação genética GENESYS, a densidade de marcadores moleculares (MM) no mapeamento de QTL (*Quantitative Trait Loci*). Foram considerados mapeamentos de alta, média e baixa resolução, em que os marcadores foram dispostos regulamente a cada 5 cM (192 MM), 10 cM (95 MM) e 20 cM (47 MM), respectivamente. Para cada intervalo, o mesmo número de marcadores utilizados foi reconsiderado, contudo, admitindo espaçamento aleatório. Os mapas foram comparados aos pares, com o mesmo número de marcadores distribuídos em intervalos regulares e de forma aleatória. A seleção assistida por marcadores foi utilizada para avaliar o desempenho fenotípico em cada cenário. Em todos os casos, os mapas que admitiram marcadores dispersos em intervalos fixos foram superiores em relação aos ganhos fenotípicos, em especial para o mapeamento de baixa resolução. Infere-se que a disposição estratégica de marcadores moleculares no mapeamento genômico é tanto mais relevante quanto menor for o número de marcadores considerados na análise.

PALAVRAS-CHAVE: Densidade de marcadores. Mapeamento. QTL. Simulação.

INTRODUÇÃO

A disponibilidade de marcadores moleculares tem possibilitado pesquisas relacionadas à genética quantitativa, tais como a identificação do número de genes afetando um determinado caráter, sua localização e o percentual da variância explicada por eles. São vários os marcadores envolvidos com características quantitativas (SAGHAI-MAROOF et al., 1996; DUDLEY, 1997).

O mapeamento de QTL (*Quantitative Trait Loci*) alicerça-se no propósito de inferir sobre os genótipos de um determinado QTL, com a finalidade de estimar seus efeitos e localizações a partir de associações com marcadores genéticos conhecidos (LIU, 1998). O mapeamento de QTL pode ser feito através da varredura genômica, utilizando marcadores genéticos espalhados por todo o genoma, identificando regiões que afetam características quantitativas. Quanto maior o número de marcadores utilizados na análise de ligação maior a resolução do mapeamento (maior densidade). Adotando métodos estatísticos adequados é possível identificar QTL, bem como estimar suas posições e seus efeitos, por meio de associações entre marcadores e características quantitativas de interesse.

Desta forma, o mapa genético ou mapa de ligação de uma determinada espécie animal ou vegetal representa um modelo abstrato do arranjo

linear de um grupo de genes e locos marcadores, ou seja, de segmentos de DNA de herança simples, que podem ser acompanhados através das gerações (LIU, 1998). Ele permite o estudo da arquitetura genética de características complexas, identificando, mapeando e mensurando a magnitude dos efeitos dos principais fatores genéticos envolvidos no controle destes caracteres (FERREIRA; GRATTAPAGLIA, 1998).

Neste contexto, a manipulação da molécula de DNA possibilita o conhecimento da localização, função e expressão de diferentes genes e, conseqüentemente, a compreensão da base genética da diversidade fenotípica das espécies animais e vegetais (ANDERSSON; GEORGES, 2004).

O número de marcadores utilizados e o tamanho da população avaliada são fatores relevantes para obtenção de associações significativas entre marcador e QTL. A utilização de poucos marcadores distribuídos aleatoriamente diminui as chances de encontrar desequilíbrio de ligação entre marcador e QTL, que é fundamental para detectar ligações entre marcadores e locos quantitativos. Já o uso de muitos marcadores genotipados e avaliados sob baixo rigor estatístico levará a falsas ligações (QTL fantasmas).

Apesar do aumento no número de estudos de mapeamento de QTL, a resolução com que os mesmos são conduzidos ainda é considerada baixa, em geral estando na ordem de 20 a 30 Centimorgan - cM (LEDUR et al., 2003). Estudos de simulação

evidenciaram que 5 cm seria uma distância adequada entre os marcadores que contêm um QTL para uso na seleção assistida por marcadores (MOREAU et al., 1998). Segundo Lande e Thompson (1990), a maior parte da variação dentro de pequenas regiões cromossômicas será devido a mudanças de uma única base (SNPs – *Simple Nucleotide Polymorphism*).

O mapeamento de alta resolução contribuirá para aumentar a eficiência da seleção assistida por marcadores e de programas de introgressão, visto que, a região genômica a ser controlada será menor (DEKKERS; HOSPITAL, 2002). Para obtenção de marcadores suficientes em número e informação que permita sua ligação com o QTL se faz necessário considerar inicialmente um maior número de marcadores estrategicamente dispostos no genoma (VAN DER BEEK et al., 1995).

Objetivou-se neste trabalho avaliar, via simulação de dados, diferentes densidades de marcadores distribuídos em intervalos regulares e de forma aleatória no mapeamento genômico, via a seleção assistida por marcadores, com relação ao desempenho fenotípico obtido ao longo de 20 gerações sob seleção.

MATERIAL E MÉTODOS

Os dados utilizados neste trabalho foram simulados via o programa computacional de simulação genética GENESYS (EUCLYDES, 2007). Este programa é escrito na linguagem de programação FORTRAN. Ele permite a criação de genomas de certa complexidade, que podem ser utilizados para formação de populações de acordo com a estrutura desejada, sob influência de questionamentos propostos a serem analisados, sejam através de métodos de seleção, pressuposições estatísticas, sistemas de acasalamentos, entre outros parâmetros.

A introdução de algoritmos de simulação genética permite avaliar populações submetidas a condições diversas, no que diz respeito a sua estrutura genômica, possibilitando uma análise simultânea de parâmetros estatísticos e biológicos. Com o desenvolvimento de aplicativos na área de genética e melhoramento o pesquisador poderá realizar uma série de análises indispensáveis a sua pesquisa, através da condensação de seus dados, sem a perda de informações. As metodologias executadas foram:

1. Simulação do genoma

Para o presente estudo um genoma hipotético foi simulado possibilitando avaliar a distribuição de marcadores ao longo do genoma,

admitindo diferentes densidades, com variações no número e no espaçamento entre os marcadores. As analogias foram feitas através dos desempenhos fenotípicos e das freqüências de alelos relacionados à característica quantitativa considerada, em 20 gerações consecutivas sob seleção.

O genoma foi caracterizado geneticamente:

- apresentava 958 cM de extensão;
- 40 cromossomos de tamanho aleatório;
- 100 locos quantitativos (QTL);
- 47, 95 e 192 marcadores moleculares distribuídos de forma regular e aleatória ao longo do genoma, ao admitir os intervalos de um marcador a cada 20, 10 e 5 cM, respectivamente;
- os efeitos aditivos dos locos quantitativos foram simulados seguindo a distribuição normal dos dados fenotípicos;
- os locos quantitativos foram dialélicos e não possuíam desvios de dominância e nem epistasia;
- não possuíam cromossomo sexual e as freqüências gênicas iniciais foram iguais para ambos os sexos;
- as freqüências gênicas iniciais para os locos quantitativos e para os marcadores moleculares seguiram distribuição normal, apresentando valores próximos a 0,5;
- os efeitos de ambiente foram simulados conforme a distribuição normal;
- os dados fenotípicos simulados apresentaram média de 2,00 unidades e desvio padrão 0,25 unidades.

2. Simulação das populações

Para a estrutura genômica simulada foi construída uma população base composta de 500 machos e 500 fêmeas, não aparentados entre si. Os 500 descendentes foram obtidos do cruzamento aleatório entre dez machos e 100 fêmeas (dez fêmeas/macho), produzindo cinco descendentes/fêmea/macho (500 indivíduos), que formavam a população inicial / experimental disponível. Esta população foi submetida à seleção assistida por marcadores durante 20 gerações consecutivas com dez repetições cada qual, visando minimizar os erros da oscilação genética.

Foi realizado um total de seis processos seletivos, diferenciando-se em relação às quantidades de marcadores admitidas e sua distribuição em intervalos regulares ou aleatoriamente.

3. Seleção Assistida por Marcadores (Marker Assisted Selection – MAS)

Foram admitidas três diferentes quantidades de marcadores de acordo com o intervalo preestabelecido, a saber: um marcador a cada 5 cM,

10 cM e 20 cM, o que representou um total de 192, 95 e 47 marcadores moleculares (MM), respectivamente, dispostos de forma estratégica ao longo do genoma. Para cada intervalo utilizou-se novamente o mesmo número de marcadores, contudo, admitindo uma distribuição aleatória, o que totalizou seis MAS, possibilitando a comparação de mapas com a mesma resolução, através da analogia entre as médias das frequências de alelos relacionados ao caráter e dos desempenhos fenotípicos.

A partir da população inicial selecionaram-se dez machos e 100 fêmeas (dez fêmeas/macho), que foram acasalados ao acaso produzindo 500 descendentes (cinco descendentes/fêmea/macho), mantendo-se o mesmo número de descendentes a cada nova geração subsequente, respeitando o delineamento estabelecido na geração zero (população inicial).

Desta forma, após a identificação de marcadores moleculares relacionados à característica de interesse, através da análise de QTL, conduziu-se a testes de DNA nas progênies resultantes dos cruzamentos, selecionando os indivíduos que possuam alelos favoráveis em homozigose para os QTL identificados. A este procedimento convencionou-se chamar de seleção

assistida por marcadores (*Marker Assisted Selection* – MAS). De acordo com esta seleção, os reprodutores foram selecionados com base nos genótipos de um número de marcadores moleculares que estariam estatisticamente associados a locos quantitativos. A identificação da associação entre marcadores e QTL foi feita através de uma análise de regressão linear simples entre os genótipos dos marcadores e os valores fenotípicos das progênies oriundas dos acasalamentos, adotando um nível de significância de 5%.

RESULTADOS

Os processos de seleção partiram do mesmo valor fenotípico na população inicial (geração 0), que foi de 2,00 unidades, permitindo a analogia da distribuição dos marcadores, ora em intervalos regulares, ora dispostos aleatoriamente, de acordo com o número admitido.

A seleção visa aumentar o valor fenotípico, objetivo alcançado em maior magnitude ao adotar a distribuição regular dos marcadores em intervalos fixos, para todas as três resoluções consideradas (baixa, média e alta) (Tabela 1).

Tabela 1. Desempenho fenotípico médio obtido ao longo de 20 gerações sob seleção assistida por marcadores, utilizando marcadores distribuídos aleatoriamente e em intervalos fixos.

Gerações	192 Marcadores		95 Marcadores		47 Marcadores	
	5 cM	Aleatórios	10 cM	Aleatórios	20 cM	Aleatórios
0	2.000	2.000	2.000	2.000	2.000	2.000
1	2.128	2.135	2.123	2.066	2.073	2.028
2	2.245	2.229	2.165	2.137	2.139	2.057
3	2.356	2.311	2.248	2.206	2.208	2.147
4	2.432	2.381	2.335	2.287	2.274	2.227
5	2.525	2.456	2.397	2.342	2.341	2.297
6	2.595	2.540	2.476	2.411	2.418	2.372
7	2.667	2.600	2.536	2.466	2.457	2.423
8	2.733	2.655	2.604	2.492	2.535	2.456
9	2.760	2.702	2.643	2.552	2.596	2.495
10	2.810	2.734	2.706	2.600	2.657	2.516
11	2.844	2.766	2.746	2.648	2.711	2.563
12	2.889	2.805	2.794	2.676	2.758	2.574
13	2.930	2.830	2.841	2.726	2.789	2.611
14	2.963	2.871	2.880	2.753	2.842	2.639
15	2.999	2.900	2.917	2.782	2.875	2.641
16	3.029	2.925	2.959	2.805	2.882	2.654
17	3.050	2.967	2.980	2.835	2.915	2.670
18	3.071	2.984	3.003	2.852	2.919	2.671
19	3.102	3.011	3.028	2.884	2.937	2.682
20	3.119	3.028	3.047	2.915	2.946	2.707
Ganhos (%)*	55,95	51,40	52,35	45,75	47,30	35,35
Diferença**	4,55%		6,6%		11,95%	

* Ganhos percentuais obtidos ao término das 20 gerações sob MAS, de acordo com a resolução do mapeamento e a distribuição dos marcadores; ** Diferença no ganho fenotípico percentual obtido ao término das 20 gerações sob MAS, entre a disposição estratégica e aleatória para cada quantidade de marcadores considerada.

A distinção no desempenho fenotípico resultante ao término das 20 gerações entre a distribuição aleatória e por intervalos é mais expressiva para o mapeamento de baixa resolução (47 MM), na qual o intervalo de 20 cM apresentou uma superioridade percentual de aproximadamente 12% (47,3% – 35,35%) em relação à disposição aleatória.

Também é válido ressaltar que a média da frequência de alelos relacionados à característica mostra-se de maior magnitude ao adotar intervalos regulares entre marcadores adjacentes, independente da resolução (Tabela 2). Uma

superioridade na frequência de alelos favoráveis ao caráter é observada para mapas de alta e média resolução, em especial ao admitir espaçamentos fixos entre os marcadores. Para mapas de baixa resolução a disposição estratégica dos marcadores em intervalos de 20 cM torna-se de maior relevância, face a maior magnitude na diferença média da frequência de alelos favoráveis entre a distribuição aleatória e a estrategicamente disposta, em comparação a distinção obtida para os mapeamentos que utilizaram 95 e 192 marcadores moleculares

Tabela 2. Frequência média de alelos relacionados à característica obtida ao longo de 20 gerações sob seleção assistida por marcadores, utilizando marcadores distribuídos aleatoriamente e em intervalos fixos.

Gerações	192 Marcadores		95 Marcadores		47 Marcadores	
	5 cM	Aleatórios	10 cM	Aleatórios	20 cM	Aleatórios
0	0.50	0.50	0.50	0.50	0.50	0.50
1	0.52	0.52	0.53	0.51	0.53	0.51
2	0.54	0.54	0.54	0.53	0.54	0.51
3	0.56	0.55	0.56	0.54	0.56	0.52
4	0.58	0.57	0.59	0.56	0.58	0.53
5	0.60	0.59	0.60	0.57	0.58	0.55
6	0.61	0.61	0.61	0.58	0.59	0.56
7	0.63	0.62	0.62	0.59	0.61	0.57
8	0.64	0.63	0.64	0.60	0.62	0.58
9	0.65	0.64	0.65	0.61	0.63	0.58
10	0.65	0.65	0.66	0.62	0.64	0.59
11	0.66	0.65	0.67	0.63	0.65	0.59
12	0.67	0.67	0.68	0.64	0.66	0.60
13	0.68	0.68	0.68	0.66	0.68	0.60
14	0.70	0.69	0.69	0.66	0.68	0.61
15	0.70	0.69	0.69	0.67	0.69	0.61
16	0.71	0.70	0.70	0.68	0.69	0.61
17	0.72	0.71	0.71	0.68	0.70	0.61
18	0.72	0.71	0.72	0.69	0.70	0.62
19	0.73	0.72	0.73	0.69	0.70	0.62
20	0.73	0.72	0.73	0.70	0.71	0.62
Média*	0.65	0.64	0.65	0.62	0.64	0.58
Diferença**		0.01		0.03		0.06

* Média da frequência de alelos relacionados à característica durante 20 gerações sob MAS, de acordo com a resolução do mapeamento e a distribuição dos marcadores; ** Diferença na média da frequência de alelos relacionados à característica entre a disposição estratégica e aleatória para cada quantidade de marcadores considerada.

DISCUSSÃO

A disponibilidade de um grande número de marcadores estrategicamente dispostos no mapeamento genômico é de grande importância na detecção de QTL, possibilitando a cobertura de todas as regiões do genoma, em todos os cromossomos. Logo, uma otimização no desempenho fenotípico poderá ser obtida em programas de seleção que utilizam mapas de alta resolução (DARVASI et al., 1993; LEE, 1995).

Os resultados evidenciaram um maior desempenho fenotípico ao praticar a MAS adotando

alta resolução e intervalos regulares entre os marcadores adjacentes no genoma. Quanto maior o número de marcadores estrategicamente distribuídos forem utilizados na busca por QTL, maior será a probabilidade de identificarem ligações significativas entre marcadores e QTL, conforme afirma Hillel (1997). Mesmo admitindo quantidades distintas de marcadores, sua disposição estratégica leva vantagem, conforme a superação no ganho percentual fenotípico ao término das 20 gerações, ao comparar a disposição de 95 marcadores moleculares (MM) em intervalos de 10 cM com 192 MM distribuídos aleatoriamente, bem como os 47

MM em intervalos de 20 cM com os 95 MM dispostos de forma aleatória.

A disposição de marcadores adotando intervalos fixos possibilita uma melhor varredura genômica, cobrindo todo o segmento de DNA, além de requerer um menor número de marcadores para determinado poder de detecção de QTL / desempenho fenotípico na MAS. A análise de QTL em intervalos de média (10 cM) e baixa (20 cM) resolução representa apenas um ponto inicial na busca por genes relacionados à característica, devido os mesmos poderem apresentar diversos genes, minimizando a precisão na detecção de QTL. Posteriormente a esta análise inicial deve-se proceder a um mapeamento mais fino para identificar marcadores mais próximos aos genes de interesse (MONFORTE; TANKSLEY, 2000).

Em contrapartida, a distribuição de marcadores aleatoriamente favorece desuniformidade na varredura genômica, visto que, existe a possibilidade de grupos de marcadores estarem muito próximos, resultando em efeito único, ou seja, dois ou mais marcadores muito próximos funcionarão como único marcador. Além disso, grandes extensões do genoma ou do cromossomo poderão não ser cobertas por marcadores, inviabilizando a identificação de suas ligações com QTL e sua utilização nas gerações subsequentes sob seleção.

A distribuição estratégica de marcadores torna-se mais relevante em mapeamentos de baixa resolução (20 cM ou 47 MM), pois maior diferenciação no ganho percentual fenotípico foi obtida entre as duas distribuições consideradas (11,95%). Já para os mapeamentos de média e alta resolução, essa diferença foi de aproximadamente 6,6% (52,35% – 45,75%) e 4,55% (55,95% – 51,40%), respectivamente.

As diferenças médias entre as frequências de alelos relacionados à característica foram de 0.06, 0.03 e 0.01, para mapeamentos de baixa, média e alta resolução, respectivamente. Os resultados confirmam a importância de admitir

intervalos regulares entre os marcadores adjacentes à medida que se utiliza um menor número de marcadores no mapeamento.

Na análise do genoma para mapeamento de QTL é aceitável que os procedimentos que utilizam marcadores espalhados em intervalos regulares otimizam a identificação de regiões que afetam caracteres quantitativos, em especial, nos mapeamentos de baixa resolução, onde se utilizam poucos marcadores moleculares para a detecção de QTL.

De acordo com Lee (1995), a eficiência da seleção assistida por marcadores além de ser otimizada pela disposição estratégica dos marcadores, poderá também ser acrescida pela ocorrência de forte ligação entre marcadores e QTL, conseguida ao admitir mapas de alta resolução (≤ 5 cM).

CONCLUSÕES

A distribuição de marcadores em intervalos regulares apresentou maior eficiência nos progressos fenotípicos obtidos ao longo das gerações sob seleção assistida por marcadores, em analogia a distribuição aleatória, de acordo com a densidade admitida no mapeamento. Esta superioridade também é observada na frequência média de alelos relacionados à característica quantitativa considerada.

Infer-se que esta relevância da distribuição estratégica de marcadores no desempenho fenotípico é mais proeminente na análise de mapas de menor resolução, ou seja, no mapeamento que utiliza pequenas quantidades de marcadores moleculares.

AGRADECIMENTOS

À Universidade Federal de Viçosa - UFV e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq, pelo apoio financeiro.

ABSTRACT: The objective of this study was to evaluate, through the program of genetic simulation GENESYS, the density of molecular markers (MM) in the mapping of QTL (*Quantitative Trait Loci*). High, medium and low resolution mapping were considered, in which markers were arranged regularly every 5 cM (192 MM), 10 cM (95 MM) and 20 cM (47 MM), respectively. For each gap, the same number of markers used was reconsidered, however, assuming random spacing. The maps were compared in pairs, with the same number of markers distributed at regular intervals and at random. The selection assisted by markers was used to assess the phenotypic performance in each scenario. In all cases, the maps which admitted sparse markers at fixed intervals were superior in relation to phenotypic gains, especially for the mapping of low resolution. We conclude that the strategic distribution of molecular markers in genomic mapping is more relevant as the lower is the number of markers considered in the analysis.

KEYWORDS: Density of markers. Mapping. QTL. Simulation.

REFERÊNCIAS

- ANDERSSON, L.; GEORGES, M. Domestic-animal genomics: deciphering the genetics of complex traits. **Nature Reviews Genetics**, Hampshire, v. 5, p. 202-212, March. 2004.
- DARVASI, A.; WEINERB, A.; MINKE, V.; WELLER, J.; SOLLER, M. Detecting marker-QTL linkage and estimating QTL gene effect and map location using a saturated genetic map. **Genetics**, Maryland, v. 134, n. 3, p. 943-951, July. 1993.
- DEKKERS, J. C. M; HOSPITAL, F. The use of molecular genetics in the improvement of agricultural populations. **Nature Reviews Genetics**, Hampshire, v. 3, n. 1, p. 22-32, Jan. 2002.
- DUDLEY, J. W. Quantitative genetics and plant breeding. **Advances in Agronomy**, New York, v. 59, p. 1-23, 1997.
- EUCLYDES, R. F. **Genesys**: Sistema de Simulação Genética. Versão 9.1, Viçosa: Fundação Arthur Bernardes – UFV. 2007.
- FERREIRA, M. E.; GRATTAPAGLIA, D. **Introdução ao uso de marcadores moleculares em análise genética**. 3.ed. Brasília: EMBRAPA/CENARGEM, 1998. 220p.
- HILLEL, J. Map-based quantitative trait loci identification. **Poultry Science**, Champaign, v. 76, p. 1115-1120, 1997.
- LANDE, R.; THOMPSON, R. Efficiency of marker-assisted selection in the improvement of quantitative traits. **Genetics**, Maryland, v. 124, n. 3, p.743-756, March. 1990.
- LEDUR, M. C.; BERTANI, G. R.; NONES, K. Genômica nos programas de melhoramento genético avícola. In: CONFERÊNCIA DE CIÊNCIA E TECNOLOGIA AVÍCOLA, Painele 3: Aplicações da Biotecnologia na Avicultura, 2003, Campinas. **Anais....** Organização: FACTA - Fundação APINCO de Ciência e Tecnologia Avícolas, 2003. p. 87-105.
- LEE, M. DNA markers and plant breeding programs. **Advances in Agronomy**, New York, v.55, p.265-344, 1995.
- LIU, B. H. **Statistical genomics**: linkage, mapping and QTL analysis. Boca Raton: CRC Press, 1998. 611p.
- MONFORTE, A. J.; TANKSLEY, S. D. Fine mapping of a quantitative trait locus (QTL) from *Lycopersicon hirsutum* chromosome 1 affecting fruit characteristics and agronomic traits: breaking lineage among QTLs affecting different traits and dissection of heterosis for yield. **Theoretical and Applied Genetics**, Berlin, v. 100, n. 3-4, p. 471-479, Fev. 2000.
- MOREAU, L.; CHARCOSSET, A.; HOSPITAL, F.; GALLAIS, A. Marker-assisted selection efficiency in populations of finite size. **Genetics**, Maryland, v. 148, p. 1353-1365, March. 1998.
- SAGHAI-MAROOF, M. A.; YUE, Y. G.; XIANG, Z. X.; STROMBERG, E. L.; RUFENER, G. K. Identification of quantitative trait loci controlling resistance to gray leaf spot disease in maize. **Theoretical and Applied Genetics**, Berlin, v. 93, n. 4, p. 539-546, 1996.
- VAN DER BECK, S.; VAN ARENDONK, J. A. M.; GROEN, A. F. Power of two-and three-generation QTL mapping experiments in an outbred population containing full-sib or half-sib families. **Theoretical and Applied Genetics**, Berlin, v. 91, n. 6-7 p. 1115-1124, 1995.