

Avaliação de diferentes níveis de significância na identificação de marcadores moleculares no melhoramento genômico¹

Marcelo Jangarelli²
Ricardo Frederico Euclides³
Ana Paula Oliveira Nogueira⁴

RESUMO

Os avanços em biologia molecular e bioinformática consolidaram a genômica no estudo e sequenciamento dos genes de um indivíduo. Objetivou-se neste trabalho avaliar diferentes níveis de significância na identificação de marcadores moleculares relacionados com características quantitativas de baixa, média e alta herdabilidades. Uma comparação entre os níveis de 1, 5, 10 e 20% foi realizada por meio do sistema computacional de simulação genética (GENESYS), utilizado para a simulação de três genomas, cada qual constituído de um único caráter quantitativo de baixa, média e alta herdabilidade, respectivamente. A partir de cada um dos genomas, simulou-se uma população base e, posteriormente, uma população inicial, procedendo à avaliação dos quatro níveis de significância, por meio da seleção assistida por marcadores moleculares, por intermédio dos seguintes parâmetros: valor fenotípico, número de marcadores usados na seleção e número de marcadores fixados na seleção. Os resultados indicaram superioridade dos níveis de significância de maior magnitude (10 e 20%) em relação aos de menores valores (1 e 5%) quanto aos valores fenotípicos e ao número de marcadores identificados e utilizados na seleção, bem como sua menor fixação para as três características, embora de forma mais expressiva para o caráter de baixa herdabilidade. Mesmo os níveis que apresentaram maior precisão na detecção de marcadores ligados aos QTLs (*quantitative trait loci*) deixaram a desejar no que diz respeito às melhorias genéticas, resultando em ganhos fenotípicos inferiores.

Palavras-chave: Herdabilidade, seleção, significância genômica, simulação, QTL.

ABSTRACT

Evaluation of different significance levels in the identification of molecular markers in genomic improvement

Advances in molecular biology and bioinformatics consolidated genomic in the study and sequencing of individual genes. The aim of this work was to evaluate different significance levels in the identification of molecular markers related to quantitative characteristics of low, medium and high heritability. A comparison among the levels of 1%, 5%, 10% and 20% was carried out by a genetic simulation software (GENESYS), used to simulate three genomes, each one consisting of only one quantitative character of low, medium and high heritability. From each genome, a base population was simulated and, afterward, an initial population, proceeding to the evaluation of the four significance levels,

Recebido para publicação em março de 2008 e aprovado em abril de 2009

¹ Extraído da dissertação de mestrado em Genética e Melhoramento pela Universidade Federal de Viçosa do primeiro autor.

² Bolsista do CNPq-Brasil (PGI), doutorando em Genética e Melhoramento do Departamento de Zootecnia da Universidade Federal de Viçosa. Rua Pio X, 11, BNH, 29902-090 Linhares - ES. E-mail: gmejanga@hotmail.com

³ Doutor, Professor Adjunto do Departamento de Zootecnia da Universidade Federal de Viçosa. Departamento de Zootecnia, Av. P H Rolfs, s/n, Campus Universitário - 36570-000 Viçosa - MG. E-mail: rbaja@ufv.br

⁴ Bolsista do CNPq-Brasil, doutoranda em Genética e Melhoramento do Departamento de Fitotecnia da Universidade Federal de Viçosa. Programa de Pós-Graduação em Genética e Melhoramento, Sala 12, Ed. Artur Bernardes, Campus UFV, 36570-000 Viçosa - MG. E-mail: anap812004@yahoo.com.br

through the selection assisted by molecular markers, using the following parameters: phenotypic value, number of used markers and fixed markers in the selection. The results indicated superiority of the significance levels of higher magnitude (10% and 20%) compared with the lower values (1% and 5%), concerning the phenotypic value obtained after selection, besides a higher number of markers identified and used in the selection, as well as a lower setting, for all the three characteristics, although in more expressively for the low heritability character. Even the levels which presented higher precision in detecting markers related to QTLs (*quantitative trait loci*) need to be developed when related to genetic improvement, resulting in lower phenotypic gain.

Key words: Heritability; selection; genomic significance; simulation; QTL.

INTRODUÇÃO

O recente desenvolvimento de algumas tecnologias, como a obtenção de marcadores moleculares cobrindo todo o genoma, a construção de mapas genéticos, o desenvolvimento e a automação de métodos de sequenciamento de DNA e o aperfeiçoamento da bioinformática, otimizaram o melhoramento genético molecular e o desenvolvimento da ciência genômica (Borém *et al.*, 2003).

Os marcadores moleculares são sequências de DNA e podem ser identificados e mapeados, sendo utilizados para identificar e localizar locos específicos nos cromossomos (Anderson *et al.*, 1994). A seleção assistida por marcadores moleculares (SAMM) é condicionada pela significância genômica exigida nas associações entre os genótipos dos marcadores e os QTLs (*quantitative trait loci*), que são regiões cromossômicas relacionadas com a variação fenotípica das características quantitativas (Hayward *et al.*, 1994). O uso dos marcadores em modernos programas de melhoramento baseia-se no princípio de que se um gene ou conjunto de genes está estreitamente associado a um marcador molecular de fácil identificação, então a seleção para esse marcador será mais eficiente do que para a própria característica (Hayward *et al.*, 1994). A SAMM consiste de dois passos principais: i) identificação de associações entre locos marcadores e QTLs e ii) uso dessas associações para o desenvolvimento de populações melhoradas (Bulfield, 1997). Entretanto, a SAMM deve ser encarada como uma ferramenta auxiliar e não substituta dos métodos tradicionais de melhoramento (Lande & Thompson, 1990).

A capacidade de detecção de associações entre marcadores e QTLs é limitada pelo tamanho da população de mapeamento, pela exigência estatística para detecção dos locos (nível de significância genômico), pelo número de marcadores utilizados no mapeamento e pela herdabilidade da característica em estudo (Liu, 1998).

Como poucos locos economicamente importantes são conhecidos e os QTLs não têm suas bases moleculares definidas, esses atualmente têm sido referidos mais como uma associação estatística do que como uma entidade

biológica, sendo a herança acompanhada pelos marcadores (Guimarães, 2004). Portanto, os QTLs têm sido identificados como associações estatísticas entre dados relativos a uma região genômica e a variabilidade fenotípica existente entre populações segregantes (Li, 1998). Por serem assim caracterizados, os procedimentos para sua localização e estimação são tipicamente baseados em testes estatísticos (Burrow & Blake, 1998), identificando marcadores mais fortemente associados com o surgimento de determinado fenótipo (William, 1998).

Objetivou-se com este trabalho avaliar diferentes níveis de significância na identificação de marcadores moleculares relacionados com características quantitativas de baixa, média e alta herdabilidades.

MATERIAL E MÉTODOS

Os dados utilizados neste trabalho foram simulados via sistema computacional de simulação genética – GENESYS (versão 2007), desenvolvido por Euclides (1996). Esse sistema foi escrito na linguagem do programação FORTRAN, permitindo a criação de genomas de certa complexidade, que podem ser utilizados para formação de populações de acordo com a estrutura desejada, com influência de questionamentos propostos a serem analisados, sejam por meio de métodos de seleção, pressuposições estatísticas, sistemas de acasalamentos, dentre outros fatores, dispensando, portanto, animais e laboratórios.

Para o presente estudo, foram simulados genomas semelhantes ao da espécie *Gallus gallus* (frango de corte), possibilitando estudar os efeitos de diferentes níveis de significância, admitidos na SAMM, em características de baixa, média e alta herdabilidades sobre os seguintes parâmetros: valor fenotípico, número de marcadores usados na seleção e número de marcadores fixados.

Utilizando-se o programa GENESYS, foram simulados três genomas, cada qual contendo 80 marcadores moleculares, sendo cada um constituído de uma única característica quantitativa com herdabilidade de 0,10, 0,35 ou 0,60, respectivamente, designada simplesmente de peso (kg), governada por 200 locos quantitativos (QTLs).

Todos os genomas simulados foram caracterizados geneticamente: com 958 centimorgans de extensão; 40 cromossomos de tamanho aleatório; os efeitos aditivos dos locos quantitativos foram simulados seguindo a distribuição normal dos dados fenotípicos; os locos quantitativos foram dialélicos e não possuíram desvios de dominância e nem epistasia; não possuíram cromossomo sexual e as frequências gênicas iniciais foram iguais para ambos os sexos; as frequências gênicas iniciais para os locos quantitativos seguiram distribuição normal, apresentando valores próximos de 0,5; as frequências gênicas iniciais para os marcadores moleculares seguiram distribuição normal, apresentando valores próximos de 0,5; os efeitos de ambiente foram simulados conforme a distribuição normal; os dados fenotípicos simulados apresentaram média de 2,00 kg e desvio-padrão de 0,20 kg.

Para cada estrutura genômica simulada foi construída uma população base composta de 500 machos e 500 fêmeas, não aparentados entre si. Os 1.000 descendentes foram obtidos do cruzamento aleatório entre 10 machos e 100 fêmeas (10 fêmeas/macho), produzindo 10 descendentes/fêmea/macho (1.000 indivíduos). Assim, formaram-se as populações iniciais que, por sua vez, foram submetidas à seleção assistida por marcadores moleculares por 20 gerações consecutivas, com 10 repetições para cada uma, visando minimizar os efeitos da flutuação genética. A partir de cada uma dessas populações iniciais os reprodutores foram selecionados com base nos genótipos de determinado número de marcadores moleculares, que estariam estatisticamente associados aos QTLs. A cada geração, os 10 machos e as 100 fêmeas (10 fêmeas/macho) que apresentaram os melhores desempenhos foram acasalados ao acaso, produzindo 1.000 descendentes (10 por acasalamento) que, por sua vez, formaram a geração seguinte.

Dessa forma, obtida a população inicial para cada genoma, foram praticadas quatro seleções assistidas por marcadores, nas quais a distinção estava no nível de significância genômico adotado para a detecção de marcadores associados aos QTLs de interesse, todas partindo do mesmo valor fenotípico. A identificação da associação dos marcadores foi feita por meio de uma análise de regressão linear simples entre os genótipos dos marcadores e os valores fenotípicos dos descendentes dos acasalamentos.

Os níveis de significância que caracterizavam os quatro processos de SAMM, por genoma/população inicial, foram:

Nível de significância de 1% - Considerado de alta significância, detectando somente os marcadores que estiverem mais fortemente relacionados ao genótipo dos QTLs e, conseqüentemente, ao fenótipo da característica de interesse. Ele apresenta maior precisão, contudo, espera-se detectar menor número de marcadores associados à característica quantitativa.

Nível de significância de 5% - Significativo, selecionando os marcadores QTLs altamente significativos a 1%, assim como os significativos para maiores valores de significância genômica; ou seja, para 2, 3, 4 ou até 5%. Dessa forma, espera-se que um número igual ou maior de marcadores seja detectado em comparação ao nível de 1%, em função da menor precisão exigida.

Nível de significância de 10% - A associação entre os marcadores e os QTLs selecionados a esse nível foi referida como sugestiva, tornando-se indicadora de possíveis associações entre seu genótipo com as expressões fenotípicas nos indivíduos. Espera-se que maior número de marcadores seja detectado, compreendendo os altamente significativos ($\alpha = 1\%$), os significativos ($\alpha = 5\%$), e os sugestivos (α entre 5 e 10%).

Nível de significância de 20% - A associação entre os marcadores e os QTLs selecionados também foi referida como sugestiva, porém com menor precisão do que os sugestivos a 10%, devido ao maior erro a eles associado (20%). Espera-se que número igual ou superior de marcadores seja selecionado, em comparação com os níveis anteriores, por representar a soma dos que foram significativos a de 1, 5 e 10%, disponibilizando maior número de informações, apesar da menor precisão.

RESULTADOS E DISCUSSÃO

Característica de baixa herdabilidade: $h^2 = 0,10$

Na Figura 1 são apresentados os valores fenotípicos médios obtidos pela SAMM. Como mencionado, a distinção dos quatro processos de seleção executados estava no nível de significância adotado, evidenciando ser a identificação de marcadores moleculares uma questão na qual a estatística apresenta grande relevância, devido aos resultados diferenciados de acordo com a significância genômica preestabelecida.

Como pode ser observado, todos os quatro processos de seleção partiram de um mesmo valor fenotípico, referindo-se à característica peso, progredindo no decorrer das gerações. A seleção visa aumentar o peso, objetivo alcançado em maior magnitude para os níveis de 5 e 10% e, em especial, para o nível de 20%. A partir da 12ª até a 14ª geração, as técnicas seletivas apresentaram uniformidade no progresso fenotípico, pois, segundo Euclides (1996) e Carneiro (1998), ocorre redução na variância genética aditiva ao longo das gerações selecionadas, além da diminuição do número de marcadores moleculares identificados e utilizados na seleção, em virtude da fixação da maioria dos locos alélicos, conforme apresentado nas Figuras 2 e 3 (Falconer, 1981).

Para as características de baixa herdabilidade, a seleção assistida por marcadores mostra-se eficiente, em comparação à seleção somente com avaliação fenotípica

(Lande & Thompson, 1990), pois, segundo Ferreira & Grattapaglia (1998), os marcadores são ferramentas eficazes para estudos genômicos, uma vez que detectam o polimorfismo diretamente em DNA e não sofrem qualquer tipo de influência ambiental (Souza, 2001). Contudo, sua eficiência não chega a ser significativa se a característica referida expressar baixa herdabilidade.

Isso foi comprovado pelo baixo ganho fenotípico obtido ao término das 20 gerações selecionadas. Mas, de acordo com a significância genômica admitida, a diferença no ganho fenotípico médio foi representativa, com superioridade de aproximadamente 62% ao adotar o nível de 20% (ganho de peso de 0,4701 kg) em comparação ao de 1% (ganho de peso de 0,2897 kg).

É válido lembrar que, ao adotar níveis de significância de menores magnitudes (1 e 5%), o pesquisador estará trabalhando com maior precisão em seus resultados (Ferreira & Grattapaglia, 1998). Nesse caso, tem-se baixa possibilidade de erro comparativamente aos níveis de maior magnitude (10 e 20%). Contudo, ao ganhar em precisão poder-se-á estar perdendo em outros parâmetros importantes para a continuidade e progresso do melhora-

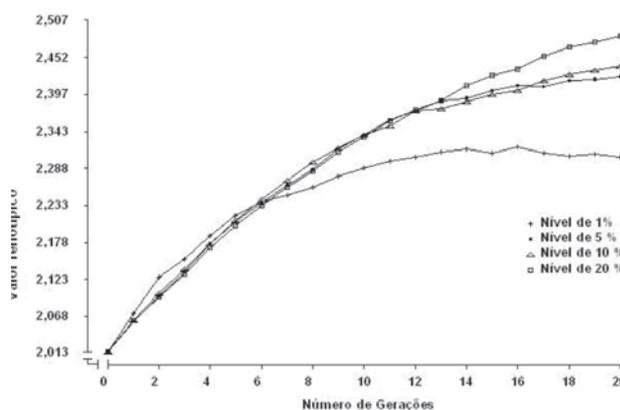


Figura 1. Valores fenotípicos médios para a característica de baixa herdabilidade durante 20 gerações sob SAMM, de acordo com o nível de significância adotado (1, 5, 10 e 20%).

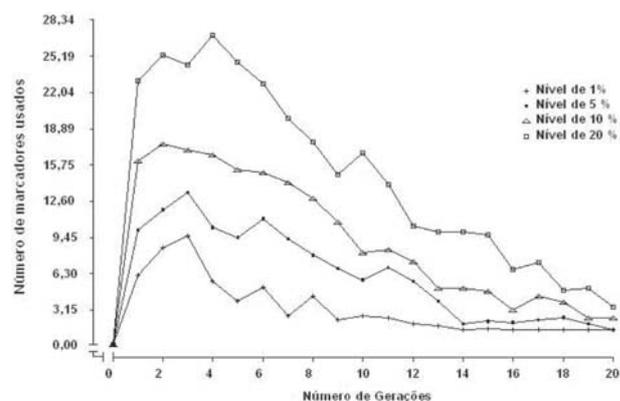


Figura 2. Número médio de marcadores usados na seleção para a característica de baixa herdabilidade durante 20 gerações sob SAMM, de acordo com o nível de significância adotado (1, 5, 10 e 20%).

mento, como o número de marcadores associados às QTLs e número de marcadores fixados na seleção.

Considera-se que a identificação da associação entre marcadores e QTLs depende do desequilíbrio de ligação entre alelos de locos marcadores e de alelos segregantes nos locos que influenciam as características fenotípicas; da proximidade entre esses diferentes tipos de locos; e da magnitude da diferença entre os efeitos genéticos dos alelos do loco quantitativo. Assim, espera-se que a SAMM que utiliza maior número de marcadores selecionados apresente, em média, melhores respostas fenotípicas e genotípicas que envolvam locos marcadores (Euclides, 1996).

Maior número de marcadores identificados e usados na SAMM foi obtido para os níveis de 10 e 20%, em virtude do efeito acumulativo. Nota-se na Figura 2 que o número de marcadores se reduz no decorrer das gerações para os quatro processos seletivos, como consequência da fixação dos marcadores que ocorre no decorrer das gerações, conforme apresentado na Figura 3 (Carneiro, 1998). A fixação foi superior para os níveis de 1 e 5% devido à exigência de grande efeito entre o genótipo do marcador e a expressão fenotípica, assim como do número restrito de marcadores selecionados, favorecendo uma uniformidade entre os marcadores em relação ao seu genótipo e, conseqüentemente, o uso mais exaustivo desses poucos marcadores selecionados.

As características quantitativas são governadas por grande número de genes, cada qual com pequeno efeito, em que a expressão do caráter é resultante do efeito coletivo de todos eles. A detecção de marcadores utilizando níveis altamente significativos poderá prejudicar o efeito coletivo dos genes relacionados ao caráter. Nesse caso, somente associações significativas entre marcadores e QTLs serão selecionadas, desprezando-se as de menor significância. Ao se adotarem níveis sugestivos serão selecionados os marcadores de maior efeito, assim como os que apresentam associação sugestiva sobre a expressão fenotípica, que, quando considerados coletivamente, trariam resultados benéficos para o caráter (Figura 1).

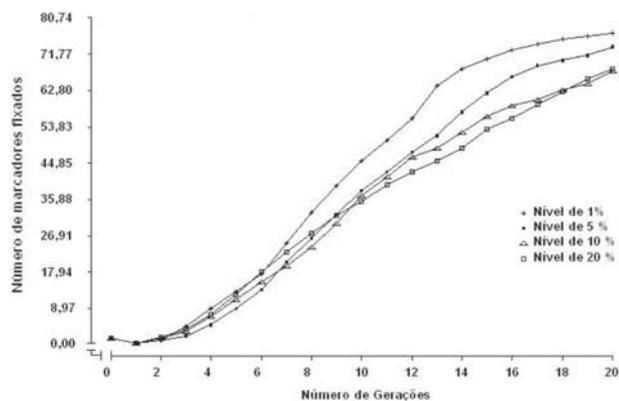


Figura 3. Número médio de marcadores fixados para a característica de baixa herdabilidade durante 20 gerações sob SAMM, de acordo com o nível de significância adotado (1, 5, 10 e 20%).

Característica de média herdabilidade:
 $h^2 = 0,35$

Novamente, maior valor fenotípico foi alcançado para os níveis de 5, 10 e 20% de significância, conforme obtido para a característica anterior. Contudo, em virtude do caráter ser de média herdabilidade, as distinções entre os processos seletivos que adotaram significâncias de 1 e 20% tornaram-se menos pronunciadas. A SAMM é de menor relevância para essa característica, pois, de acordo com Lande & Thompson (1990), quanto menor a herdabilidade de um caráter maior será a eficiência da seleção assistida por marcadores.

Para a presente característica, a diferença no ganho fenotípico ao término das 20 gerações, entre os dois limites extremos de significância admitidos, foi de aproximadamente 30%, inferior à obtida para o mesmo caractere, adotando uma herdabilidade de magnitude menor ($h^2 = 0,10$).

De forma semelhante, maior número de marcadores é detectado ao se adotar os níveis de 10 e 20% (Figura 5). Também se observa redução no número de marcadores

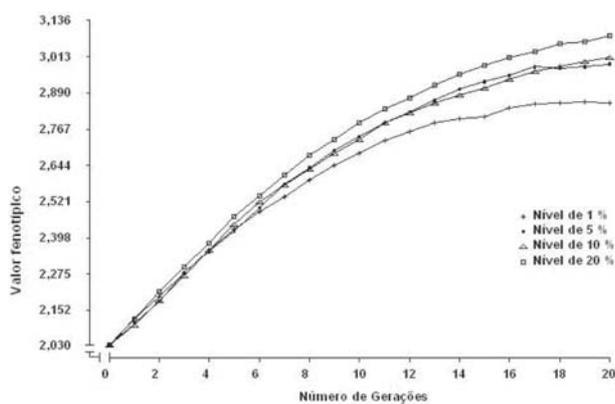


Figura 4. Valores fenotípicos médios para a característica de média herdabilidade durante 20 gerações sob SAMM, de acordo com o nível de significância adotado (1, 5, 10 e 20%).

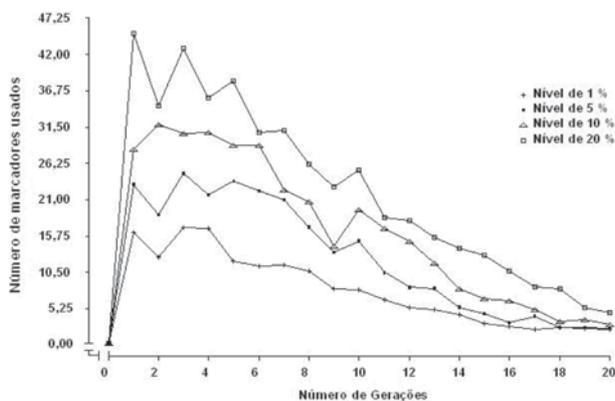


Figura 5. Número médio de marcadores usados na seleção para a característica de média herdabilidade durante 20 gerações sob SAMM, de acordo com o nível de significância adotado (1, 5, 10 e 20%).

detectados ao longo das gerações, como consequência da fixação de locos marcadores (Figura 6).

Maior número de marcadores ligados aos QTLs é identificado e utilizado nesta característica, devido ao fato de o caráter apresentar herdabilidade maior ($h^2 = 0,35$), em analogia à anterior ($h^2 = 0,10$), tornando a expressão fenotípica mais precisa e confiável quanto aos vínculos genotípicos.

Característica de alta herdabilidade:
 $h^2 = 0,60$

Para essa característica, a amplitude na distinção entre os níveis de significância não foi tão expressiva quanto à apresentada para as características de baixa e média herdabilidade, face à menor eficiência da SAMM e, consequentemente, à menor relevância da associação estatística na identificação dos marcadores ligados aos QTLs. Apesar disso, os níveis sugestivos contribuíram para obtenção de melhores resultados fenotípicos, com superioridade de aproximadamente 10% do nível de maior magnitude (20%) sobre o de menor valor (1%) ao término das 20 gerações.

Semelhantemente às características anteriores, que se diferenciavam apenas na magnitude da herança do caráter, ao se adotarem os níveis sugestivos evidencia-se maior número de marcadores detectados e menor fixação de locos marcadores (Figuras 8 e 9, respectivamente). O número de marcadores usados foi ainda maior, em comparação com os caracteres anteriores, pois a expressão fenotípica do caráter apresenta alto vínculo com seu genótipo.

A seleção assistida por marcadores é de maior relevância para caracteres com herdabilidade de baixa magnitude, na qual maiores distinções foram observadas entre os diferentes níveis de significância admitidos. Em características de alta herdabilidade a expressão fenotípica dos indivíduos é um indicativo do seu valor reprodutivo, resultando em maiores valores fenotípicos a serem obser-

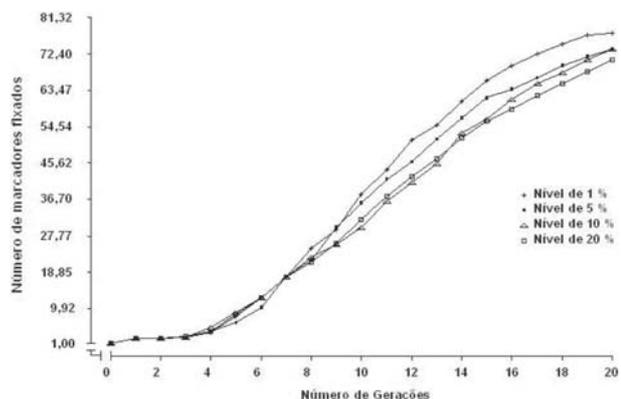


Figura 6. Número médio de marcadores fixados para a característica de média herdabilidade durante 20 gerações sob SAMM, de acordo com o nível de significância adotado (1, 5, 10 e 20%).

vados nos processos seletivos. Além disso, essa maior confiabilidade entre o valor fenotípico e o genético do indivíduo propicia maior eficiência na detecção de marcadores associados aos QTLs.

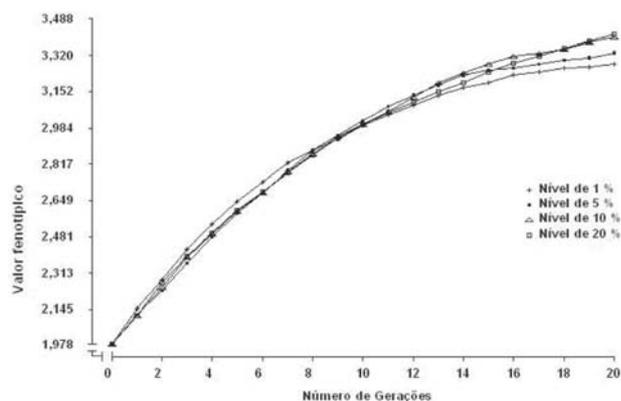


Figura 7. Valores fenotípicos médios para a característica de alta herdabilidade durante 20 gerações sob SAMM, de acordo com o nível de significância adotado (1, 5, 10 e 20%).

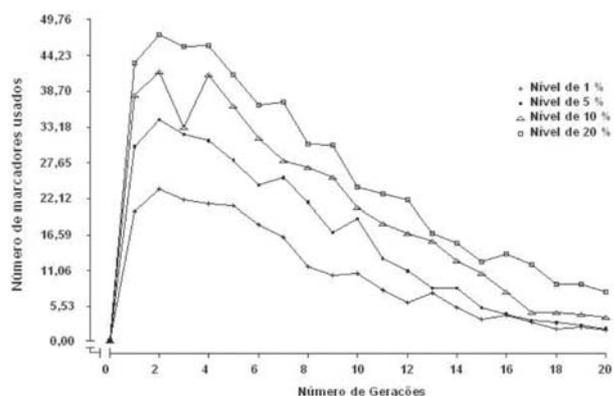


Figura 8. Número médio de marcadores usados na seleção para a característica de alta herdabilidade durante 20 gerações sob SAMM, de acordo com o nível de significância adotado (1, 5, 10 e 20%).

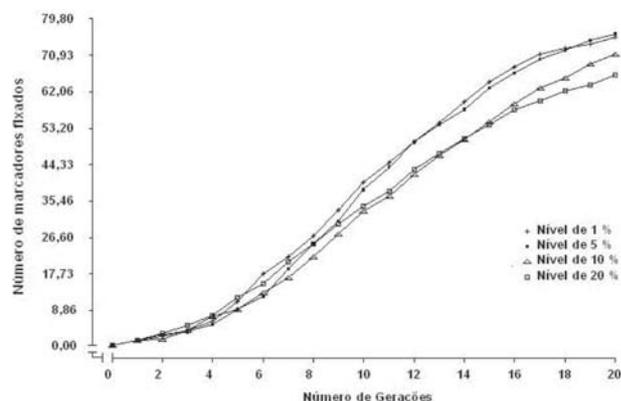


Figura 9. Número médio de marcadores fixados para a característica de alta herdabilidade durante 20 gerações sob SAMM, de acordo com o nível de significância adotado (1, 5, 10 e 20%).

CONCLUSÕES

Adotando diferentes níveis de significância na SAMM obtiveram-se ganhos fenotípicos distintos como resposta à seleção. Os níveis de 1 e 5% propiciaram menores ganhos fenotípicos e os de 10 e 20% os melhores resultados. Ao se relacionar a herdabilidade da característica, os resultados foram otimizados para o caráter em que a herança era de magnitude menor.

Os níveis de alta significância (1 e 5%) proporcionaram detectar menor número de marcadores, minimizando as respostas a serem alcançadas com a seleção assistida por marcadores moleculares.

REFERÊNCIAS

- Anderson L, Haley CS, Ellegren H, Knott SA, Johansson M, Anderson K, Anderson-Eklund L, Edfors-Lilja I, Fredholm M, Hansson I, Hakansson J & Lundstrom K (1994) Genetic mapping of quantitative trait loci for growth and fatness in pigs. *Science*, 263:1771-1774.
- Borém A, Giúdice M & Sedyama T (2003) *Melhoramento Genômico*, Viçosa, UFV. 224 p.
- Bulfield G (1997) Strategies for the future. *Poultry Science*, 76:1071-1074.
- Burrow MD & Blake TK (1998) Molecular tools for the study of complex traits. In: *Molecular dissection of complex traits*. New York: 1ª ed. CRC. p. 13-30.
- Carneiro PLS (1998) Efeito de erros de pedigree na avaliação genética animal. Dissertação de mestrado. Viçosa, Universidade Federal de Viçosa. 79 p.
- Euclides RF (1996) Uso do sistema para simulação Genesys na avaliação de métodos de seleção clássicos e associados a marcadores moleculares. Tese de doutorado. Viçosa, Universidade Federal de Viçosa. 149 p.
- Falconer DS (1981) *Introdução a genética quantitativa*, Viçosa, UFV. 279 p.
- Ferreira ME & Grattapaglia D (1998) *Introdução ao Uso de Marcadores Moleculares em Análise Genética*, 3ª edição. Brasília, Embrapa Cenargen. 220 p.
- Guimarães SEF (2004) Análise de marcadores genômicos e detecção de QTLs e genes candidatos em melhoramento animal. In: *Pereira JCC (4.ed.) Melhoramento Genético Aplicado a Produção Animal*. Belo Horizonte, FEPMVZ Editora. p. 491-524.
- Hayward MD, Mcadam NJ, Jones JG, Evans C, Evans GM, Forster JW, Ustin A, Hossain KG, Quader B, Stammers M & Will JK (1994) Genetic markers and the selection of quantitative traits in forage grasses. *Euphytica*, 77:269-75.
- Lande R & Thompson R (1990) Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics*, 124:743-756.
- Li Z (1998) Molecular analysis of epistasis. In: *Molecular dissection of complex traits*. New York: 1ª ed. CRC. p. 119 - 130.
- Liu BH (1998) *Statistical genomics: linkage, mapping and QTL analysis*, Boca Raton, CRC Press. 611 p.
- Souza AP (2001) *Biologia molecular aplicada ao melhoramento*. In: *Nass LL, Valois ACC, Mello IS & Valadares-Ingliš MC (Ed.) Recursos genéticos e melhoramento de plantas*. Rondonópolis, Fundação MT. p. 939-965.
- William DB (1998) QTL analysis: power, precision and accuracy. In: *Paterson AH Molecular dissection of complex traits*. New York, CRC Press. p. 145-162.