46ª Reunião Anual da Sociedade Brasileira de Zootecnia
Maringá, PR - UEM - 14 a 17 de julho de 2009

Análise de Associação por Todo o Genoma para Identificar Locos Relacionados ao Lucro Líquido, à Vida Produtiva e ao Escore de Células Somáticas na raça Jersey[1]

Marcos V. B. Silva[1,2], Paul M. VanRaden[3], Curt P. Van Tassell[1], Tad S. Sonstegard[1], Lakshmi K. Matukumalli[4], Eui-Soo Kim[1], George R . Wiggans[3]

[1]Bovine Functional Genomics Laboratory, Agricultural Research Service, USDA Beltsville, MD, 20705, EUA. Email: curtvt@ars.usda.gov; tads@ars.usda.gov; EuiSoo.Kim@ars.usda.gov;
[2]Embrapa Gado de Leite, Juiz de Fora, MG, 36038-330, Brasil – Bolsista do CNPq. E-mail: marcos@cnpgl.embrapa.br
[3]Animal Improvement Programs Laboratory, Agricultural Research Service, USDA Beltsville, MD, EUA. Email: Paul.VanRaden@ars.usda.gov; George.Wiggans@ars.usda.gov
[4]Bioinformatics and Computational Biology, George Mason University, Manassas, VA 20110, EUA. Email: Lakshmi.Matukumalli@ars.usda.gov

Resumo: Foi realizada uma varredura por todo o genoma de animais da raça Jersey, nos EUA, utilizando marcadores do tipo SNP, visando identificar QTL associados ao lucro líquido, à vida produtiva e ao escore de células somáticas. Os dados usados neste estudo foram provenientes do *Animal Improvement Programs Laboratory*, USDA/EUA. Amostras de DNA coletadas de 2.380 animais da raça Jersey e as habilidades preditas de transmissão de 2.081 animais, publicadas em fevereiro/2009 (http://greenbook.usjersey.com/), foram utilizadas nas análises. Para a genotipagem dos SNPs foi usado o BovineSNP50 BeadChip da Illumina, com aproximadamente 54.000 SNPs. Todo SNP com *call rate* (<99%), em desequilíbrio de Hardy-Weinberg (teste exato p<0,01) e com frequência de um dos alelos menor do que 5% foram excluídos das análises finais (30.342 SNP usados). P-values corrigidos pelo teste de Bonferroni iguais a 0,01 foram usados. Para todas as características, SNP significativos foram encontrados em vários cromossomos, especialmente nos BTAs 3, 4, 5, 6, 10, 12, 23 e 25. Os resultados sugerem que as soluções para os efeitos dos marcadores nas avaliações genômicas podem identificar regiões cromossômicas que necessitem ser melhor estudadas.

Palavras–chave: associação por todo genoma, bovino, escore de célula somática, lucro líquido, SNP, vida produtiva.

Genome-Wide Association Analysis to Identify Loci for Net Merit, Productive Life, and Somatic Cell Score in Jersey Breed

Abstract: A genome scan was conducted in the US Jersey population to identify QTL affecting net merit, productive life, and somatic cell score. Data used at this study were traditional genetic evaluations computed from DHI records by Animal Improvement Programs Laboratory, USDA/EUA. DNA has been acquired for 2,380 animals from Jersey breed and genetic evaluations released in Feb/2009 (http://greenbook.usjersey.com/) for 2,081 animals were used for the whole-genome association. Genotyping was performed using Illumina's BovineSNP50 BeadChip with approximately 54,000 SNPs. For GWA study, any SNP with low call rate (<99%), departure from Hardy-Weinberg equilibrium (exact test p<0.01), and minor allele frequency below 5 percent, were excluded from the final analysis (30,342 markers retained). Bonferroni corrected level of 0.01 were used. For all traits, significant SNPs were found on many chromosomes, especially on BTAs 3, 4, 5, 6, 10, 12, 23, and 25. Most parts of the significant SNPs were located in similar positions on the chromosomes for all traits. Some of them localized within a single linkage disequilibrium (LD) block. Results suggest that marker solutions from genomic evaluations may be useful for identifying chromosomal regions that merit further study.

Keywords: bovine, net merit, productive life, SNP, somatic cell score, whole-genome association.

Introdução

Até recentemente, o número de marcadores de DNA identificados no genoma de bovinos era limitado e o custo de genotipagem era alto, restringindo os experimentos delineados para o mapeamento de QTL. Mesmo com o uso de mapas de ligação esparsos, baseados em microssatélites, avanços substanciais têm sido obtidos na identificação de regiões cromossômicas que contenham locos que estejam associados a características de importância econômica em animais domésticos. Em bovinos de leite, QTL associados à produção de gordura e de proteína e/ou produção de leite foram mapeados por Viitalla et al (2003) e Szyda et al. (2005).

Após o desenvolvimento do BovineSNP50 BeadChip da Illumina, com aproximadamente 54.000 SNPs (Van Tassell et al., 2007), tornou-se possível a genotipagem de bovinos em larga escala, sendo possível cobrir todo o genoma com marcadores com a distância média de 0,05 cM.

O lucro líquido (LL), considerado como índice chave na seleção de touros, a vida produtiva (VP) e o escore de céluas somáticas (ECS), que são fatores impactantes na redução dos custos nos sistemas de produção de leite, são características importantes no melhoramento genético de gado de leite. Assim, objetivou-se com este estudo, usar mapas densos de SNP para a identificação das regiões do genoma que estão associadas com essas características em animais da raça Jersey, criados nos EUA.

Material e Métodos

Os dados usados neste estudo são provenientes do *Animal Improvement Programs Laboratory*, USDA (Beltsville, MD, EUA). Amostras de DNA coletadas de 2.380 animais da raça Jersey e as habilidades preditas de transmissão para o LL, a VP e o ECS de 2.081 animais, publicadas em fevereiro/2009 *(*http://greenbook.usjersey.com/*)*, foram utilizadas para as análises de associação de todo o genoma. Os genótipos dos marcadores SNP foram obtidos por meio do BovineSNP50 BeadChip (Illumina, San Diego,CA, EUA). Para a maioria dos SNPs, os genótipos foram lidos com menos do que 0,1% de taxa de erros. Foram excluídos da análise os SNPs redundantes, aqueles com frequência de um dos alelos menor do que 0.05 e em desequilíbrio de Hardy-Weinberg (teste exato $p<0.001$). Após a edição do arquivo de genótipos, o número de marcadores SNP caiu de 50.386 para 30.342.

As análises estatísticas foram realizadas por meio de programas escritos na linguagem Fortran e por scripts em ambiente R. P-values foram corrigidos para os testes em todo o genoma e maximização dos modelos genéticos e nível de significância em todo o genoma igual a 0,05 (bi-caudal). P-values iguais a 0,01, corrigidos pelo teste de Bonferroni, foram usados. Cinco modelos genéticos (co-dominante, dominante, recessivo, sobredominante e log-aditivo) foram analisados.

Foram selecionados os cinco SNPs com maior P-value no modelo log-aditivo (SNPs candidatos) para cada característica e, por meio das respectivas coordenadas genômicas (*assembly* BTAU_4.0), foi feita a anotação dos genes mais próximos utilizando-se o Genome Browser (http://genome.ucsc.edu/).

Resultados e Discussão

Os resultados da varredura de todo o genoma, para as três características estudadas, são apresentados na Figura 1. As análises de marcadores únicos identificou vários sinais que excederam o limiar de 5×10^{-7}, particularmente nos cromossomos 3, 4, 5, 6, 10, 12, 23 e 25. Tais sinais demarcam regiões do genoma onde possivelmente estejam localizados genes que tenham grande influência sobre o LL, a VP e o ECS.

Na Tabela 1 são apresentados os cinco SNPs mais significativos para as três características, suas coordenadas genômicas e os genes mais próximos a eles. Constatou-se que quatro SNPs são os mesmos para LL e VP, sendo ARS-BFGL-NGS-17920 o marcador com o sinal mais forte para ambas. Uma das explicações para esse resultado pode ser a forte associação genética entre essas duas características, como foi evidenciado por VanRaden (2004), o qual encontrou valor de 0,58 para a correlação genética entre elas, na raça Holandesa. Ainda, deve ser destacado o fato de que, a partir de 1994, a VP foi combinada com outras características de produção para a obtenção do índice LL. Ressalte-se, ainda, que o gene PTHLH está relacionado à alterações da glândula mamária (Jonas et al., 2008).

Para o ECS, foram evidenciadas fortes associações no BTA3 e no BTA23, especialmente para SNPs próximos aos locos da Catepsina Z (CTSZ), do Complexo Principal de Histocompatibilidade Bovina (BoLA) e do Fator de Tumor Necrose α (TNF-α), os quais estão relacionados à regulação do sistema imunológico.

Deve ser ressaltado que as associações não provêem identificação precisa dos genes causais, tampouco da mutação causativa. Para tal fim, seriam necessários esforços para o mapeamento fino e o re-sequenciamento da região de interesse, seguidos de estudos funcionais que possam traduzir tais inferências em afirmações mais robustas acerca dos mecanismos moleculares e fisiológicos envolvidos.

Conclusões

Foram encontradas associações de marcadores e regiões no BTA5, centradas no SNP ARS-BFGL-NGS-17920, as quais tiveram grandes efeitos sobre o lucro líquido e a vida produtiva. Resultados sugerem que as soluções para os efeitos dos marcadores nas avaliações genômicas podem identificar regiões cromossômicas que necessitem ser melhor estudadas.

Literatura citada

Jonas, E.; Schreinemachers, H-J. ; Kleinwächter, T., et al. QTL for the heritable inverted teat defect in pigs. **Mammalian Genome**, v.19, p. 127-138, 2008.

Szyda, J., Liu, Z., Reinhardt, F., Reents, R. Estimation of quantitative trait loci parameters for milk production traits in German Holstein dairy cattle population. **Journal of Dairy Science**, v. 88:356–367, 2005.

Van Tassell, C.P., Matukumalli, L., Taylor, et al. 2007. **Construction and application of a bovine high-density SNP assay**. American Dairy Science and Society of Animal Science. San Antonio, USA.(CD-ROM)

VanRaden, P.M. 2004. Selection on Net Merit to Improve Lifetime Profit. **Journal of Dairy Science**, v. 87, p.3125–3131.

Viitala, S. M., Schulman, N. F., de Koning, D. J., et al. 2003. Quantitative trait loci affecting milk production traits in Finnish Ayrshire dairy cattle. **Journal of Dairy Science,** v. 86, p. 1828–1836.

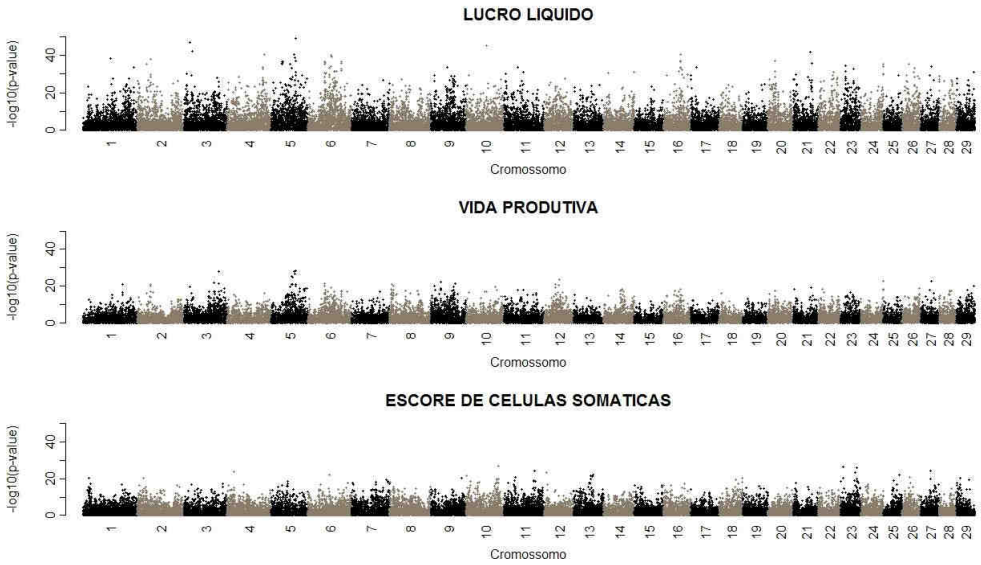

Figura 1 Varredura do genoma bovino para as características estudadas. Cromossomos são mostrados em cores alternadas (preto e cinza) para melhor visualização.

Tabela 1 Regiões do genoma apresentando fortes sinais de associação para cada característica estudada

Característica	BTA	Genes	SNP	Posição	P-value
Lucro Líquido	5	CASC1	ARS-BFGL-NGS-17920	91062496	$7{,}07 \times 10^{-50}$
	3	TNFAIP8L2	BTA-66810-no-rs	19967708	$6{,}38 \times 10^{-48}$
	5	FAR2	ARS-BFGL-NGS-98512	86679675	$4{,}40 \times 10^{-41}$
	5	PTHLH	Hapmap43491-BTA-122155	88671798	$2{,}37 \times 10^{-39}$
	5	CASC1	BTA-74438-no-rs	91247447	$6{,}56 \times 10^{-38}$
Vida Produtiva	5	CASC1	ARS-BFGL-NGS-17920	91062496	$4{,}80 \times 10^{-29}$
	3	TIE1	ARS-BFGL-NGS-70648	109793953	$1{,}39 \times 10^{-28}$
	5	FAR2	ARS-BFGL-NGS-98512	86679675	$1{,}42 \times 10^{-28}$
	5	PTHLH	Hapmap43491-BTA-122155	88671798	$2{,}56 \times 10^{-27}$
	5	TIMP3	ARS-BFGL-NGS-25351	76985632	$6{,}13 \times 10^{-26}$
Escore de Células Somáticas	13	CTSZ	BTA-32994-no-rs	58010284	$1{,}97 \times 10^{-22}$
	23	BoLA	Hapmap46836-BTA-55820	28043180	$4{,}04 \times 10^{-12}$
	23	BoLA-DRB3	BTA-56042-no-rs	26442665	$5{,}07 \times 10^{-12}$
	23	TNF-α	ARS-BFGL-NGS-16619	28007940	$8{,}97 \times 10^{-12}$
	23	BoLA	ARS-BFGL-NGS-90570	28140135	$5{,}52 \times 10^{-11}$