

CONTROLE DE QUALIDADE DE GENÓTIPOS E AMOSTRAS UTILIZADOS NA SELEÇÃO GENÔMICA DAS RAÇAS BRAFORD E HEREFORD

COMIN, Helena Brocardo¹; SOLLERO, Bruna Pena²; JUNQUEIRA, Vinicius Silva³; CARDOSO, Leandro Lunardini²; HIGA, Roberto Hiroshi⁴; CAETANO, Alexandre Rodrigues⁵; YOKOO, Marcos Jun-Iti⁶; CARDOSO, Fernando Flores⁶

Palavras-Chave: Bovinos. Genotipagem. Heterozigocidade.

Introdução

A incidência de carrapato bovino *Rhipicephalus (Boophilus) microplus* na região dos trópicos gera perdas produtivas e reprodutivas, especialmente naqueles animais mais susceptíveis ao parasita, afetando expressivamente a economia nos sistemas de produção da pecuária bovina.

Com os avanços da metodologia de seleção genômica, e, portanto, das aplicações de estratégias envolvendo genotipagens em diferentes escalas, o significativo incremento no ganho genético anual em características de alto impacto econômico vem se tornando uma realidade. No entanto, avaliar anteriormente, a qualidade de marcadores e amostras por meio de alguns critérios de controle de qualidade (CQ), é passo crítico para garantir que as análises estatísticas sejam mais fidedignas (Yang et al., 2011). Além disso, com a otimização daqueles SNP's e amostras mais representativas, pretende-se diminuir a carga computacional sem acarretar em perda de acurácia, podendo ainda auxiliar na convergência das soluções para os efeitos dos marcadores e valores genômicos calculados a partir destes efeitos (Wiggans et al., 2009).

Este estudo teve como propósito, avaliar diferentes critérios de controle de qualidade em marcadores genéticos do tipo polimorfismo de base única (SNP – *Single Nucleotide Polymorphism*) e amostras que posteriormente serão utilizados nas predições de diferenças esperadas na progênie (DEP) genômicas para resistência ao carrapato bovino nas raças Braford e Hereford.

¹Estudante de graduação em Zootecnia, UNIPAMPA, FAPERGS, helenacomin.92@hotmail.com

²Pós-doutoranda (o), Embrapa Pecuária Sul, CAPES, brunasollero@yahoo.com.br, lunardini.cardoso@ufrgs.br

³Mestrando em Genética e Melhoramento, UFV, CAPES, junqueiravinicius@hotmail.com

⁴Pesquisador A, Embrapa Informática Agropecuária, roberto@cnptia.embrapa.br

⁵Pesquisador A, Embrapa Recursos Genéticos e Biotecnologia, acaetano@cenargen.embrapa.br

⁶Pesquisador A, Embrapa Pecuária Sul, Embrapa, fernando.cardoso@cppsul.embrapa.br, marcos.yokoo@embrapa.br

Metodologia

Os animais das raças Hereford e Braford utilizados no presente trabalho (N=3.681) foram amostrados de quatro safras de rebanhos pertencentes ao programa de melhoramento genético da associação de criadores Conexão Delta G, criados no Rio Grande do Sul. Amostras de sangue, pelo ou sêmen dos indivíduos foram coletadas para a extração do DNA, e seus dados de genealogia, performance e contagem de carrapatos foram registrados. Além de 3.551 animais genotipados com o painel *Illumina BovineSNP50* (54.609 SNPs), outros 130 touros (pais) foram genotipados com o painel de alta densidade (*BovineHD - Illumina® bead chip* – 777.962 SNPs). Para proceder ao controle de qualidade em nível de marcadores e de amostras, rotinas no programa R versão 3.0.2 (*R Development Core Team*, 2008) foram desenvolvidas utilizando-se o pacote *snpStat* (Clayton, D., 2012). Dentre os critérios de CQ e seus limites de exclusão para os SNPs, avaliou-se o *Call rate* (CR), que é a proporção de SNP's identificados (<98%); *Minor Allele Frequency* (MAF), em que se determina a frequência mínima dos SNPs (<3%); existência do Equilíbrio de Hardy-Weinberg (HWE), de acordo com as frequências alélicas, podendo identificar possíveis erros de genotipagem ($P=10e-7$); o nível de correlação entre os SNPs ($r > 0,98$) e SNPs na mesma posição. Para as amostras, os critérios e limites foram: CR (<90%); desvio de heterozigocidade acima de três desvios padrões (>3DP) em relação à média; genótipos idênticos (>99,5%) e erros de identificação de sexo, no caso de indivíduos identificados como machos apresentarem genótipos heterozigotos para marcadores no cromossomo "x". Os critérios de controle de qualidade foram aplicados aos genótipos retornados a partir do painel de 50K, e posteriormente aos genótipos advindos do painel de alta densidade (HD). Finalmente, ambos os resultados foram confrontados para quantificar o número final de marcadores (em comum entre os dois painéis de marcadores) e indivíduos a serem utilizados em avaliações genômicas.

Resultados e Discussões

Por meio das análises de CQ das genotipagens advindas do painel de 50K (Tabela 1), do total de 52.886 marcadores (autossomais) testados, foram excluídos 5.157 SNPs por obterem CR inferior a 98%, 1.900 SNPs devido a desvios significativos ($P=1e-6$) em relação HWE e 6.444 SNPs devido ao critério MAF. Segundo Wiggans et al. (2009) e VanRaden et al. (2009) o critério de MAF deveria variar entre 2-5%, dependendo do número de amostras a serem genotipadas.

Quanto maior o número de amostras, menor a porcentagem de MAF estabelecida. Com base em outros critérios, 31 SNP's, no presente trabalho, foram descartados por apresentarem alta correlação com outros SNP's, e 27, por estarem localizados na mesma posição cromossômica que outros. Estes dois últimos critérios são também importantes para identificar aqueles SNPs mais ou menos representativos. Réplicas de SNPs e/ou genótipos pouco divergentes entre amostras podem ocorrer devido a problemas de qualidade da genotipagem, inclusive. Para o controle de qualidade de amostras genotipadas pelo painel de 50K, foram excluídos 33 animais que apresentaram taxa de heterozigocidade $> 3DP$, 49 de acordo com o critério CR (considerados como amostras de DNA de baixa qualidade), oito por apresentarem mais do que 99,5% de genótipos idênticos, e oito devido a erros de identificação de sexo. Vale ressaltar que apenas os cromossomos autossômicos e SNPs com posição conhecida foram utilizados no QC, à exceção da inclusão do cromossomo sexual "x" para a análise de averiguação do sexo do animal. Como apresentado na tabela 2, um total de 41.045 marcadores SNPs foram selecionados pelo controle de qualidade (78%), bem como 3.461 amostras (98%).

Paralelamente, os mesmos critérios foram aplicados para os resultados dos 130 animais genotipados por meio do painel de alta densidade (Tabela 1). Neste caso, a única amostra sugerida a ser excluída, por apresentar taxa de heterozigocidade $< 3DP$, era referente a um animal híbrido (Nelore x Hereford), ou seja, da geração F1 na formação do Braford, portanto, com uma expectativa de taxa de heterozigocidade muito maior que os animais Braford de geração 3/8 avançada ou Hereford; optando-se por considerá-la nas análises subsequentes. Em relação aos marcadores, do total de 735.293 SNP's, 642.234 (87%) foram aprovados pelos critérios de controle de qualidade (Tabela 2).

Comparando-se os dois painéis de marcadores após CQ, detectou-se 34.032 SNPs em comum, sendo que 7.013 outros marcadores que estavam contidos no painel de 50k, mas não naquele de 777K, foram considerados como genótipos perdidos para os 130 animais analisados com o painel de alta densidade. Finalmente, 3.591 registros de animais permaneceram no arquivo de dados em que 41.045 marcadores SNPs foram considerados para as avaliações posteriores de DEP's genômicas.

Conclusão

Os critérios de controle de qualidade para filtrar os marcadores (ambos os painéis – 50k e 777k) e amostras, utilizados no presente trabalho, são capazes de gerar dados para avaliações genômicas mais acuradas para resistência a carrapatos nos animais das raças Branford e Hereford. Além disso, avaliar (em nível genômico) simultaneamente duas raças por meio deste grupo de marcadores selecionados simplifica a manipulação dos dados.

Em pesquisas futuras, pretende-se testar se a consideração das duas raças separadamente, bem como a exclusão de alguns critérios de controle de qualidade (consequentemente alterando os SNP's selecionados) afetariam a acurácia das avaliações genômicas.

Tabela 1. Contagem de amostras e SNP's descartados por critério de exclusão

Critérios/ Razão de Exclusão (Amostras)	50K	777K
<i>Call Rate</i> < 90%	49	0
Desvio Heterozigocidade > 3 SD	33	0
Genótipos Idênticos > 99.5%	8	0
Erros Identificação Sexo	8	0
Critérios/ Razão de Exclusão (SNP's)		
<i>Call Rate</i> < 98%	5157	70747
<i>Minor Allele Frequency</i> (MAF) < 3%	6444	22060
Desvio HWE (P=10e-7)	1900	2108
Correlação entre SNP's (> 0,98)	31	981
SNP's na mesma posição	27	47

Tabela 2. Contagem de amostras e SNP's testados e selecionados para avaliações genômicas.

Marcadores e Amostras testados e selecionados	50K	777K	Total
SNP's testados (autossomais)	52886	735293	
SNP's selecionados para avaliação genômica	41045	642234	41045
Amostras genotipadas	3551	130	
Amostras selecionadas para avaliação genômica	3461	130	3591

Referências

CLAYTON, D. (2012). *snpStats*: SnpMatrix and XSnpmatrix classes and methods. R package version 1.11.0.

VANRADEN, P. M.; VAN TASSELL, C. P.; WIGGANS, G. R.; SONSTEGARD, T. S.; SCHNABEL, R. D.; TAYLOR, J. F. and SCHENKEL, F. S. Invited review: Reliability of genomic predictions for North American Holstein bulls. **Journal of Dairy Science**, v. 92, p. 16–24, 2009.

WIGGANS, G.R.; SONSTEGARD, T. S.; VANRADEN, P. M.; MATUKUMALLI, L. K.; SCHNABEL, R. D.; TAYLOR, J. F.; SCHENKEL, F. S. and VAN TASSELL, C. P. Selection of single-nucleotide polymorphisms and quality of genotypes used in genomic evaluation of dairy cattle in the United States and Canada. **Journal of Dairy Science**, v. 92, p. 3431–3436, 2009.

YANG, H-C; LIN, H-C, KANG, M., CHEN, M-H, LIN, C.W.; LI, L-H; WU, J-Y; CHEN, Y-T and PAN, W-H. SAQC: SNP Array Quality Control. **BMC Bioinformatics**, v.12, p. 1-14, 2011.