

Análises preliminares de controle de qualidade em um banco de dados de SNP genotipados em alta densidade para futuros estudos de assinaturas de seleção em bovinos da raça Canchim

Ismael Urbinati¹

Fabiana Barichello Mokry²

Danísio Prado Munari³

Roberto Hiroshi Higa⁴

Estudos com genotipagem permitem analisar regiões específicas do genoma, associando-as com o fenótipo do animal, por isso são de grande utilidade na produção animal. A genotipagem com chips de alta densidade baseia-se em marcadores genéticos de alta densidade, dentre os quais estão, atualmente, os polimorfismos de único nucleotídeo, Single Nucleotide Polymorphism (SNP), que são mudanças de um único par de bases do DNA e que permitem a identificação, em cada animal, de grande número de marcadores moleculares. Além de permitir analisar regiões do genoma associadas a uma característica (estudos de associação), estudar marcadores SNP também permite identificar regiões que estão sendo mantidas de geração para geração devido à seleção. Tais regiões são denominadas assinaturas de seleção, e são detectáveis por diferentes tipos de metodologias, utilizando dados de animais genotipados.

A identificação de regiões de assinaturas de seleção em um rebanho de bovinos de corte, que apresenta grande adaptabilidade ao clima brasileiro como o Canchim, permite que se estude os genes presentes nessas regiões e sua relação com os fenótipos utilizados na seleção da população. Isto

¹ FCAV, Unesp Jaboticabal - ismael.urbinati@gmail.com

² Universidade Federal de São Carlos

³ Departamento de Ciências Exatas, FCAV Unesp Jaboticabal

⁴ Embrapa Informática Agropecuária - roberto.higa@embrapa.br

permite uma melhor compreensão dos processos biológicos envolvidos nas manifestações fenotípicas e pode fornecer subsídios para estudos futuros de seleção genômica ampla. Por isso, o estudo de regiões de assinaturas de seleção é de grande importância para o melhoramento animal e, em particular, para a raça Canchim.

A genotipagem está sujeita a erros, e devido à alta densidade de dados (> 700.000 SNP), estes erros podem alterar de forma significativa os resultados, criando vieses. Portanto, para se realizar um estudo com dados genotipados de alta densidade, inicialmente é preciso realizar um controle de qualidade nestes dados, tanto com critérios para os SNP quanto para os animais (amostras). Este trabalho tem por objetivo analisar um banco de dados de bovinos de corte da raça Canchim genotipados com painéis de alta densidade, representantes de animais sob seleção, para posteriores escaneamentos em busca de assinaturas de seleção.

Foram utilizados registros de 400 animais genotipados com o painel Illumina BovineHD BeadChip (777.692 SNP), provenientes da base de dados genômicos da Embrapa Pecuária Sudeste, São Carlos, SP. O conjunto de dados é constituído por 205 fêmeas e 195 machos, em que 192 animais são oriundos da fazenda da Embrapa Pecuária Sudeste com origem em 17 touros, sendo 184 animais Canchim e 8 animais pertencentes ao grupo genético MA (filhos de touros Charolês e vacas 1/2 Canchim X 1/2 Zebu). O restante das amostras pertence a fazendas do Estado de São Paulo (38 animais Canchim e 9 animais MA) e de Goiás (60 animais Canchim e 95 animais MA), e 6 touros (5 Canchim e 1 Charolês) que são pais de alguns dos indivíduos genotipados. O critério para escolha dos animais genotipados foi baseado em animais (machos e fêmeas) com valores genéticos extremos (altos e baixos) para a característica área de olho de lombo.

O controle de qualidade dos SNP pode ser verificado por meio da frequência mínima de alelos (MAF), das porcentagens de SNP válidos por animal e de cada SNP em toda a população (call rate) e pelo escore de qualidade de genotipagem de cada SNP (GC score). Os valores mínimos para estes critérios foram estabelecidos pela avaliação do banco de dados e baseados em revisão de literatura. Para realizar estas análises de controle de qualidade foi utilizado o pacote "snpStats" (CLAYTON, 2012) do software estatístico R (R CORE TEAM, 2012). Na revisão de literatura, em trabalhos para fins de busca de assinaturas de seleção, foi possível encontrar alguns valores de referência para início das análises até o presente momento (Tabela 1). Os

resultados após o controle de qualidade com os parâmetros citados estão na Tabela 2.

Tabela 1. Valores referência para parâmetros de controle de qualidade para os dados disponíveis e objetivos propostos.

Valores-referência	MAF	Call rate SNP	Call rate amostra
Utsunomiya et al. (20130 (CQ 1)	0,03	90 %	90 %
Somavilla (2012) (CQ 2)	0,01	95 %	90 %

Tabela 2. Resultados preliminares para CQ utilizando os parâmetros encontrados (CQ 1 e CQ 2).

Valores-referência	Nº SNP antes CQ	Nº SNP após CQ
CQ 1	742.851	597.629
CQ 2	742.851	591.465

Os parâmetros foram similares quanto ao número de SNP excluído, ficando à critério do pesquisador qual será utilizado em futuros estudos, buscando assinaturas de seleção.

Referências

CLAYTON, D. **SnpStats and XSnpmatrix classes and methods**: R package. Version 1.8.1. 2012. Disponível em: <http://rgm3.lab.nig.ac.jp/RGM/R_rdfile?f=snpStats/man/snpStats-package.Rd&d=R_BC>. Acesso em: 27 set. 2013.

R CORE TEAM. R: **A language and environment for statistical computing**. Vienna: R Foundation for Statistical Computing, 2013. Disponível em: <<http://www.R-project.org/>> Acesso em: 27 set. 2013.

SOMAVILLA, A. L. **Identificação de assinaturas de seleção e anotação do genoma em bovinos da Raça Nelore**. 2012. 64 f. Dissertação (Mestrado em Genética e Melhoramento Animal) - Faculdade de Ciências Agrárias e Veterinárias, Universidade Estadual Paulista, Jaboticabal, 2012.

UTSUNOMIYA, Y. T.; PÉREZ O'BRIEN, A. M.; SONSTEGARD, T. S.; VAN TASSELL, C. P.; do CARMO, A. S.; MÉSZÁROS, G.; SÖLKNER, J.; GARCIA, J. F. Detecting loci under recent positive selection in dairy and beef cattle by combining different genome-wide scan methods. **PLoS ONE**, v. 8, n. 5, p. e64280, May 2013. Doi:10.1371/journal.pone.0064280, 2013.