

Controle de qualidade de dados genotípicos obtidos a partir do *Equine SNP50 BeadChip* para estudos genômicos em animais de corrida e trabalho da raça Quarto de Milha¹

Natália Andrea Rincon Beltran², Camila Tangari Meira², Josineudson Augusto II Vasconcelos Silva³, Marcilio Dias Silveira da Mota³, Rogério Abdallah Curi³

¹Parte da tese de doutorado do primeiro autor, financiada pela FAPESP

²Doutorandas do Programa de Pós-Graduação em Melhoramento Genético Animal – UNESP/Jaboticabal. Bolsista da Capes. e-mail: nataliarinconbeltran@hotmail.com camilatangari@zootecnista.com.br

³Departamento Melhoramento e nutrição Animal - UNESP/Botucatu.

Resumo: Controle de qualidade de dados de genotipagem para 188 equinos da raça Quarto de Milha (120 da linhagem de corrida e 68 de trabalho) e 54.602 SNPs, genotipados utilizando-se o *Illumina Equine SNP50 BeadChip*, foram realizadas por meio do programa *Genome Studio* versão 2011.1. Animais com *call rate* <0,95, heterozigosidade com ± 3 desvios padrão da média, e com erro na estimação do gênero foram excluídos do conjunto amostral. Em adição, foram avaliadas concordâncias de quatro réplicas e das relações de parentesco entre quatro conjuntos garanhão/progênie e três conjuntos garanhão/égua/progênie. Com respeito aos SNPs, foram excluídos os localizados no cromossomo X, com baixa qualidade de genotipagem (*cluster separation* <0,3), *call frequency* <0,9, *minor allele frequency* (MAF) <0,05 e os com *p-value* <0,001 para o equilíbrio de Hardy-Weinberg. O controle de qualidade para indivíduos levou à exclusão de quatro animais de trabalho do conjunto amostral total por apresentarem *call rate* <0,95. Considerando-se os indivíduos que passaram pelo controle de qualidade, permaneceram para análises genéticas subsequentes 40.787 SNPs na linhagem de corrida e 41.232 SNPs na linhagem de trabalho.

Palavras-chave: cavalos, DNA, polimorfismos

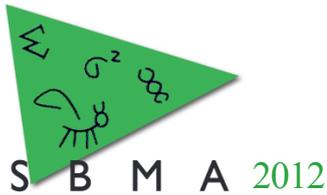
Quality control in genotype data obtained from *Equine SNP50 BeadChip* for genomic studies in animal racing and cutting Quarter Horse

Abstract: The quality control genotype data for 188 equine Quarter Horses (120 racing and 68 cutting line) and 54,602 SNPs, genotyped with the *Illumina Equine SNP50 BeadChip* was investigated using the *Genome Studio* v. 2011.1 program. Animals with *call rate* <0,95, heterozygosity ± 3 standard deviations from the mean, and errors in gender estimation were excluded. Agreement between four replicates and parentage concordance between four stallion/progeny and three stallion/mare/progeny pairs were evaluated. With respect to the quality of SNP genotypes, SNPs located on the X chromosome, with low genotyping quality (*cluster separation* <0,3), *call frequency* <0,9, MAF <0,05, and Hardy-Weinberg *p-value* <0.001 were excluded. Quality control of individual led to the exclusion of four cutting animals from the sample because of low *call rate*. Considering the individuals that passed quality control, remained for further genetic analysis 40,787 SNPs in racing and 41,232 SNPs in cutting line.

Keywords: DNA, horses, polymorphisms

Introdução

Projetado para permitir a identificação de SNPs e genes que contribuem para características de interesse nas principais raças de equinos, o *Equine SNP50 BeadChip* da empresa *Illumina* constitui-se em poderosa plataforma para estudos genéticos na espécie. Entretanto, a maior parte dos SNPs identificados em equinos, depositados em bancos de dados públicos e utilizados na confecção do *Equine SNP50 BeadChip* ainda não foi amplamente avaliada. Isto significa que o nível de polimorfismo desse conjunto de SNPs ainda é desconhecido para a quase totalidade das raças equinas criadas ao redor do mundo.



IX Simpósio Brasileiro de Melhoramento Animal

João Pessoa, PB – 20 a 22 de junho de 2012

Os objetivos do presente trabalho foram avaliar o potencial de aplicação do *Equine SNP50 BeadChip* na raça Quarto de Milha visando a possibilidade de sua aplicação em análises genéticas e programas de melhoramento.

Material e Métodos

Foram coletadas amostras de 5 mL de sangue de 188 equinos da raça Quarto de Milha, em propriedades do estado de São Paulo, sendo 120 da linhagem de corrida (18 machos e 102 fêmeas) e 68 da de trabalho (26 machos e 42 fêmeas), nascidos entre os anos de 1985 e 2009, e registrados na Associação Brasileira de Criadores de Cavalos Quarto de Milha (ABQM).

DNA genômico foi extraído das amostras de sangue utilizando-se o *Illustra Blood GenomicPrep Mini Spin Kit* (GE Healthcare, USA), de acordo com as instruções do fabricante. A genotipagem de 54,602 SNPs foi realizada com o *Illumina Equine SNP50 BeadChip* utilizando-se o equipamento HiScan (Illumina Inc., USA), junto à Faculdade de Ciências Agrárias e Veterinárias, Unesp, Jaboticabal/SP.

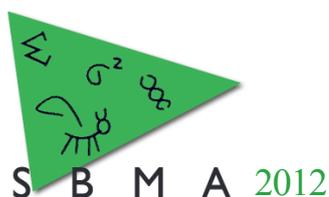
Controles de qualidade das informações de genotipagem, para indivíduos e SNPs, foram realizados por meio do programa *Genome Studio* versão 2011.1 (Illumina, USA). Em relação aos indivíduos, foram excluídos do conjunto amostral animais com *call rate* abaixo de 0,95, heterozigosidade com ± 3 desvios padrão da média, e com erro na estimação do gênero. Foram avaliadas ainda as concordâncias de quatro réplicas e das relações de parentesco (compartilhamento de alelos) entre quatro conjuntos garanhão/progênie e três conjuntos garanhão/égua/progênie. No que diz respeito aos SNPs, para cada uma das linhagens, foram excluídos (filtrados) os localizados no cromossomo X, com baixa qualidade de genotipagem (*cluster separation* inferior a 0,3), com *call frequency* inferior a 0,9, com *minor allele frequency* (MAF) inferior a 0,05 (incluindo os fixados) e com *p-value* inferior a 1×10^{-3} para o equilíbrio de Hardy-Weinberg.

Resultados e Discussão

O controle de qualidade das informações de genotipagem para indivíduos levou à exclusão de quatro animais da linhagem de trabalho do conjunto amostral total por apresentarem *call rate* $< 0,95$. As 184 amostras restantes (120 corrida e 64 trabalho) apresentaram *call rate* médio de $0,9929 \pm 0,0054$. A heterozigosidade média estimada para o total de indivíduos foi de $0,3468 \pm 0,0148$, não levando a exclusão de nenhum animal com ± 3 desvios padrão da média ($0,3023 - 0,3914$). Da mesma forma, nenhum indivíduo foi excluído por apresentar estimativa do gênero diferente do esperado. As amostras genotipadas em duplicata apresentaram concordância igual ou superior a 99,8%. Similarmente, as relações de parentesco entre os conjuntos garanhão/progênie e garanhão/égua/progênie apresentaram resultados entre 0,9985 e 0,9995.

Os SNPs mapeados no cromossomo X foram excluídos uma vez que somente são informativos em indivíduos do sexo feminino. SNPs com baixos valores de *cluster separation* (inferior a 0,3) foram removidos por apresentarem baixa precisão na diferenciação de indivíduos homozigotos para o primeiro alelo, heterozigotos e homozigotos para o segundo alelo. SNPs com *call frequency* abaixo de 0,9 foram excluídos por apresentarem perda de informação de genótipo em mais de 10% dos indivíduos estudados. SNPs com MAF inferior a 0,05 foram removidos por não serem informativos e os com *p-value* inferior a 1×10^{-3} para equilíbrio de Hardy-Weinberg foram excluídos tendo em vista que desvios extremos podem indicar problemas na precisão da genotipagem, bem como polimorfismos de ocorrência duplicada no genoma (Tabelas 1 e 2). De acordo com Ziegler et al. (2009) SNPs genotipados de forma inconsistente ou que não contribuirão para a acurácia das avaliações genéticas devem ser excluídos no sentido de reduzir esforços computacionais, diminuir a quantidade de resultados falsos e melhorar a precisão das análises realizadas com os polimorfismos restantes. A quantidade de SNPs informativos nas linhagens de corrida (40.787) e de trabalho (41.232) foi semelhante a achados disponíveis na literatura realizados com o *Equine SNP50 BeadChip* em outras raças de equinos tais como a Trotador Francês (41.249 – Teysseire et al., 2012).

As médias das MAFs dos SNPs genotipados nas linhagens de corrida ($0,2267 \pm 0,1543$) e trabalho ($0,2256 \pm 0,1496$) foram comparadas, não mostrando diferença estatística significativa (teste F; $p > 0,05$). Estes valores encontram-se próximos ao relatado por Corbin et al. (2010) em estudo de desequilíbrio de ligação em cavalos Puro Sangue Inglês ($0,30 \pm 0,12$). Entretanto, foi verificada maior



ocorrência de polimorfismos com MAF inferior a 0,05 na linhagem de corrida em relação à de trabalho (Tabela 3), o que pode ser decorrente de maior prática de acasalamentos consanguíneos e de maior intensidade de seleção praticada nesta linhagem.

Tabela 1 Número de SNPs excluídos pelo controle de qualidade dos dados de genotipagem de equinos de corrida da raça Quarto de Milha e razão para a exclusão.

Categoria/razão para a exclusão	n
SNPs genotipados	54.602
Localizados no cromossomo X	2.539
Cluster separation < 0,3	1.972
Call frequency < 0,9	143
Fixados ou com MAF < 0,05	8.925
Equilíbrio de Hardy-Weinberg – $p < 1 \times 10^{-3}$	236
SNPs selecionados para análises genéticas	40.787

Tabela 2 Número de SNPs excluídos pelo controle de qualidade dos dados de genotipagem de equinos de trabalho da raça Quarto de Milha e razão para a exclusão.

Categoria/razão para a exclusão	n
SNPs genotipados	54.602
Localizados no cromossomo X	2.539
Cluster separation < 0,3	1.972
Call frequency < 0,9	227
Fixados ou com MAF < 0,05	7.958
Equilíbrio de Hardy-Weinberg – $p < 1 \times 10^{-3}$	674
SNPs selecionados para análises genéticas	41.232

Tabela 3 Número e porcentagem de SNPs distribuídos em quatro classes de *minor allele frequency* (MAF) nas linhagens de corrida e de trabalho de animais da raça Quarto de Milha.

Linhagem	<i>Minor allele frequency</i> (MAF)			
	<0,05	0,05 – 0,5	0,05 – 0,4	0,4 – 0,5
Corrida	8.925 (17,87%)	41.023 (82,13%)	31.580 (63,22%)	9.443 (18,91%)
Trabalho	7.958 (15,96%)	41.906 (84,04%)	33.084 (66,35%)	8.822 (17,69%)

Conclusões

A taxa de SNPs informativos no genoma das linhagens de trabalho e de corrida da Quarto de Milha, indicaram que o *Equine SNP50 BeadChip* pode ser utilizado com diferentes propósitos na raças tais como na análise da estrutura genética, na estimação do grau de diversidade dentro e divergência genética entre populações, e na identificação de QTL e de regiões do genoma sujeitas à seleção.

Agradecimentos

À FAPESP e CAPES pelo apoio financeiro.

Literatura citada

- Corbin LJ, Blott SC, Swinburne JE, et al. Linkage disequilibrium and historical effective population size in the Thoroughbred horse. **Animal Genetics**, v.41, p.8-15, 2010.
- Teyssèdre S, Dupuis MC, Guérin G, Schibler L, Denoix JM, Elsen JM, et al. Genome-wide association studies for osteochondrosis in French Trotters. **Journal of Animal Science**, v.90, p.45-53, 2012.
- Ziegler A. Genome-Wide Association Studies: Quality Control and Population-Based Measures. **Genetic Epidemiology**, v.33, p.45-50, 2009.