

ESTUDIO DEL USO DE LOS MARCADORES GENÉTICOS EN UN PROGRAMA DE MEJORA GENÉTICA DE *NELORE* A TRAVÉS DE COMPARACIÓN DE MODELOS PARA LA CIRCUNFERENCIA ESCROTAL.

Rezende¹, F.M., Ibáñez-Escriche, N., Ferraz, J.B.S., Eler, J.P., Silva, R.C.G. y Mattos, E.C.

¹ Universidad de Sao Paulo. Av. Duque de Caxias Norte N° 225. 13635-900 Pirassununga, Sao Paulo, Brazil. frezende@usp.br

INTRODUCCIÓN

El desarrollo reciente de los genotipados de nucleótidos polimórficos simples (SNP) de alta densidad ha incrementado el interés por la aplicación de la selección asistida por marcadores (MAS) a escala de todo el genoma, lo que ha sido denominado selección genómica (Meuwissen et al., 2001). Sin embargo, su aplicación práctica permanece básicamente limitada a el vacuno de leche. Además, no hay todavía un consenso sobre cuál es la metodología o estrategia más adecuada a aplicar (Van Raden et al., 2008; González-Recio et al., 2008; Legarra et al., 2009). En el ganado Cebú, particularmente en la raza *Nelore*, los estudios sobre la aplicación de los marcadores genéticos se han centrado en investigar las asociaciones entre los polimorfismos individuales y caracteres cuantitativos (Ferraz et al., 2009; Pinto et al., 2010). Hasta donde sabemos, no existen trabajos científicos donde se estudie el uso de la MAS para estimar los valores genómicos con marcadores SNP en un programa de mejora genética en *Nelore*. El objetivo de este estudio es realizar una primera evaluación de la aplicación de la MAS en un programa de mejora genética de ganado de carne *Nelore*. Para llevar a cabo este objetivo se analizaron y compararon diferentes modelos de evaluación genética para el carácter circunferencia escrotal.

MATERIAL Y MÉTODOS

Datos. Los datos de este estudio pertenecen a la compañía brasileña de mejora genética de vacuno de carne Agro-Pecuaria CFM Ltda. El carácter analizado fue la circunferencia escrotal (CE), medida a los 18 meses de edad. Previo a los análisis, los datos individuales de CE fueron corregidos por los efectos sistemáticos, del grupo genético aditivo, materno y del ambiente permanente. Dichos efectos, fueron estimados con el programa VCE versión 6.0 (Groenevelt, 2008) ajustando un modelo cuatrivariante y usando la base de datos completa de este rebaño, compuesta por un pedigrí de 235.766 animales y 62.275 animales con medida de CE.

Datos genotípicos. Los genotipos utilizados correspondieron a un panel de 220 marcadores *SNP*. Las frecuencias alélicas y genotípicas de cada marcador fueron estimadas por simple conteo, usando la función *PROC FREQ* del *SAS*. Aquellos marcadores *SNP* con una frecuencia alélica menor de 5% fueron eliminados de la base de datos. Como resultado, un total de 106 marcadores *SNP* fueron usados en los análisis.

Modelos estadísticos. Tres modelos mixtos fueron usados para predecir los efectos genéticos y de los marcadores. Modelo 1, donde solamente se consideraron los efectos genéticos aditivos: $\mathbf{y} = \mu + \mathbf{Za} + \mathbf{e}$; modelo 2, donde solamente se consideraron los efectos de los marcadores: $\mathbf{y} = \mu + \mathbf{Xb} + \mathbf{e}$; modelo 3, donde se incluyeron tanto los efectos de los marcadores como los efectos genéticos aditivos: $\mathbf{y} = \mu + \mathbf{Xb} + \mathbf{Za} + \mathbf{e}$. Donde, \mathbf{y} es el vector del fenotipo corregido (1.088); μ es la media general; \mathbf{b} es el vector de los efectos de sustitución alélica de los marcadores con dimensión 106×1 ; \mathbf{a} es el vector de los efectos genéticos aditivos (7.976); \mathbf{X} es la matriz de incidencia conocida de orden $n \times p$ ($n = 1.088$; $p = 106$), cuyos elementos fueron definidos como un modelo aditivo, correspondiendo los valores 1, 2 o 3 a los genotipos aa, Aa and AA, respectivamente; \mathbf{Z} es la matriz de incidencia de los efectos genéticos aditivos de orden $n \times q$ ($n = 1.088$; $q = 7.976$); y \mathbf{e} es el vector de los efectos residuales.

Análisis estadístico. Los datos fueron analizados con inferencia Bayesiana usando métodos *Markov chain Monte Carlo* (MCMC). El programa usado en los análisis fue el TM (Legarra, et al. 2008). Distribuciones a priori impropias fueron asumidas para los parámetros de dispersión y los efectos de los marcadores, mientras que para los efectos genéticos aditivos y residuales se asumieron distribuciones normales $N(0, \mathbf{A}\sigma_a^2)$ y $N(0, \mathbf{I}\sigma_e^2)$, donde \mathbf{A}

es la matriz de relaciones de parentesco y σ_a^2 y σ_e^2 son, respectivamente, las varianzas aditiva y residual.

Comparación de modelos. Para la selección de los modelos se usaron dos criterios diferentes: *Deviance Information Criteria* (DIC) y validación cruzada k -veces. El DIC comparó la calidad global los modelo, teniendo en cuenta la complejidad de ellos (Spiegelhalter et al., 2002). Modelos con un menor DIC indican un ajuste global mejor, teniendo en cuenta su complejidad.

Dos estrategias distintas de validación cruzada k -veces se usaron para evaluar los modelos en base a su habilidad predictiva de los “datos futuros”: 1) validación cruzada 4-veces, donde la base de datos fue dividida aleatoriamente en cuatro subconjuntos distintos, cada uno conteniendo alrededor de un cuarto de los datos. En la validación cruzada, se utilizaron tres de los cuatro subconjuntos para el ajuste y estimación de los parámetros del modelo (muestra de ajuste) y la muestra de datos restante (muestra de validación) se usó para testar la habilidad predictiva del modelo; 2) validación cruzada 1-vez, la muestra de validación se compuso por los animales más jóvenes, con un total de 103 animales, los cuales representaron cerca del 10% de la base de datos.

La habilidad predictiva de los modelos fue evaluada por dos parámetros: el error cuadrático medio (MSE) y la correlación de *Pearson* (PC).

RESULTADOS Y DISCUSIÓN

Las medias posteriores de las heredabilidades (desviaciones estándar) de CE para los modelos 1 y 3 fueron 0.46 (0.09) and 0.47 (0.09), respectivamente, lo que está de acuerdo con los valores descritos por Van Melis et al. (2010; 0.42) en una muestra de mayor tamaño de la misma población de *Nelore*.

Las estimas obtenidas para el DIC fueron 1556.21, 1206.74 y 1624.04 para los modelos 1, 2 y 3, en este orden. El modelo 2, el cual incluyó solamente los efectos de los marcadores, presentó el mejor ajuste global, tal como indica su menor DIC.

La Tablas 1 y 2 muestran los resultados del MSE y la PC para los análisis de las validaciones cruzadas 4-veces y 1-vez, respectivamente. Para este panel de marcadores, los dos criterios (MSE y PC) utilizados para evaluar la capacidad predictiva presentaron el mismo patrón de resultados. El modelo 1, donde solamente se consideraron los efectos genéticos aditivos, se reveló como el modelo preferido, ya que mostró el menor valor de MSE y el mayor valor de PC para todas las validaciones cruzadas y todos subconjuntos analizados.

Los resultados obtenidos por los dos criterios de comparación de modelos fueron divergentes. El modelo 2, que sólo incluyó los efectos de los marcadores, fue el peor modelo evaluado en términos de la capacidad de predicción y el mejor modelo en términos del ajuste global (DIC). No obstante, teniendo en cuenta que el objetivo de un programa de mejora genética es predecir el rendimiento de la siguiente generación, el modelo 1 fue el mejor modelo para los dos parámetros usados en la validación cruzada MSE y PC. En conclusión, para este conjunto de marcadores y esta población, no es evidente la ventaja de incluir los efectos de los marcadores en la evaluación genética de ganado *Nelore* para CE. Por otra parte, la metodología estadística aplicada puede haber influido en los resultados. Dado que la estimación de muchos efectos con pocos datos por medio de la metodología de mínimos cuadrados puede conducir a una reducción de la correlación entre los valores observados y predichos con la disminución del número de datos (Meuwissen et al., 2001). Por tanto, son necesarios otros estudios incluyendo la aplicación de métodos de penalización (es decir, GS-BLUP, Bayes A, B, C y Bayesian Lasso), que evitan parte de ese problema.

REFERENCIAS BIBLIOGRÁFICAS

- Ferraz, J.B.S., Pinto, L.F.B., Meirelles, F.V., Eler, J.P., Rezende, F.M. et al. 2009. Genet. Mol. Res. 8 (4):1360-1366.
- González-Recio, O., Gianola, D., Long, N., Weigel, K.A., Rosa, G.J.M. et al. 2008. Genetics 178:2305–2313.
- Groeneveld, E., Kovac, M. & Mielenz, M. 2008. VCE, Institute of Farm Animal Genetics, Germany.
- Legarra, A., Aguilar, L. & Misztal, I. 2009. J. Dairy Sci. 92:4656–4663.
- Legarra, A., Varona, L. & López de Maturana, E. 2008. <http://snp.toulouse.inra.fr/~alegarra>.
- Meuwissen, T.H.E., Hayes, B.J. & Goddard, M.E. 2001. Genetics 157:1819–1829.
- Pinto, L.F.B., Ferraz, J.B.S., Meirelles, F.V., Eler, J.P., Rezende, F.M. et al. 2010. Genet. Mol. Res. 9 (3):1431-1442.
- Spiegelhalter, D.J., Best, N.G., Carlin, B.P. & Van derLinde, A. 2002. J. R. Stat. Soc. Ser. B Stat. Methodol. 64:583–639.
- Van Melis, M.H., Oliveira, H.N., Eler, J.P., Ferraz, J.B.S., Casellas, J. & Varona, L. 2010. Genet. Mol. Res. 9 (1):176-187.
- VanRaden, P. M. 2008. J. Dairy Sci. 91:4414–4423.

Tabla 1. Estimaciones de los MSE y las PC de los modelos 1, 2, 3 para los subconjuntos 1, 2, 3, 4 por la validación cruzada 4-veces.

Subconjuntos	MSE			PC		
	Modelo 1	Modelo 2	Modelo 3	Modelo 1	Modelo 2	Modelo 3
1	7.20	7.75	7.40	0.25	0.06	0.14
2	7.06	7.42	7.12	0.23	0.08	0.15
3	6.63	7.24	7.05	0.23	0.07	0.15
4	7.65	8.24	7.76	0.21	-0.02	0.08

Tabla 2. Estimaciones de los MSE y las PC de los modelos 1, 2, 3 por la validación cruzada 1-vez.

Criterio	Modelo 1	Modelo 2	Modelo 3
MSE	5.53	7.09	6.81
PC	0.32	0.01	0.13

Study of using genetic markers on a Nellore breeding program for scrotal circumference by model comparison

ABSTRACT: Data of a commercial Nellore beef cattle selection program were used to evaluate MAS implementation on its breeding program by model comparison. The data file consisted of 1,088 records for scrotal circumference (SC) adjusted for systematic, maternal and dam permanent environmental effects. A set of 106 SNP markers was analyzed to estimate the genomic breeding value. A total of three models were used to evaluate the MAS implementation on this study. Model 1 included only polygenic effects, model 2 included only markers effects and model 3 included both polygenic and markers effects. Bayesian inference via Markov chain Monte Carlo methods, performed by TM program, was used to analyze the data. Two criteria were adopted for model comparison: deviance information criteria (DIC) and k-fold cross-validation. Two k-fold cross-validation strategies were applied: 4-fold and 1-fold cross-validation. The posterior means of heritability estimated for SC from models 1 and 3 were in agreement with the estimates reported in the literature. The estimates of the DIC indicated the highest credibility of model 2. However, model 2 was the worst evaluated model in terms of prediction ability. Considering the ability to predict the next generation performance, model 1 was the best model for both the MSE and PC criterions. Therefore, it was not clear the advantage of including marker effects on the Nellore genetic evaluation for SC. Nevertheless, further studies including the application of penalized methods (i.e. GS-BLUP, Bayes A, B, C and Bayesian Lasso) are needed in order to confirm these outcomes.

Keywords: MAS, SNP markers, Zebu cattle, cross-validation.