SELECCIÓN GENÓMICA EN EL GANADO VACUNO DE LECHE ESPAÑOL

Jiménez-Montero, J.A.[1*], González-Recio, O.[2] y Alenda, R.[1]
[1]Departamento de Producción Animal. ETSI Agrónomos. UPM Ciudad Universitaria s/n. 28040 Madrid. [2]INIA. [*]Correo electrónico: joseantonio.jimenez.montero@upm.es

INTRODUCCIÓN

La selección genómica está cambiando profundamente los programas de mejora genética y el mercado del vacuno de leche. En la actualidad, es posible obtener mayor precisión en las valoraciones genéticas de animales muy jóvenes sin la necesidad del fenotipo propio o el de sus hijas. En este trabajo se estudian de las bases para la implementación de un programa de selección genómica en el vacuno de leche. Para ello se realizaron estudios de simulación y valoraciones genómicas con datos reales de la primera población nacional de referencia.
El objetivo principal fue contribuir a la implementación de la selección genómica en el vacuno de leche español. Los objetivos específicos fueron: (1) Estudiar alternativas de genotipado en poblaciones reducidas de vacuno lechero. (2) Desarrollar y validar metodología para la evaluación de grandes cantidades de genotipos. (3) Estudiar el efecto de los procesos de imputación de genotipos en la capacidad predictiva de los genotipos resultantes.

MATERIAL Y MÉTODOS

La población de referencia española estaba formada por algo más de 1,600 toros con prueba de progenie, tamaño insuficiente para obtener predicciones genómicas competitivas. Por tanto, debían evaluarse diferentes alternativas para incrementar la capacidad predictiva de las evaluaciones. Se realizó un estudio de simulación comparando estrategias de genotipado selectivo en poblaciones de hembras frente a la selección tradicional o selección genómica con una población de referencia de machos.
Posteriormente, la entrada en Eurogenomics implica el acceso a una población de referencia de más de 22,000 toros y por tanto posibilita evaluaciones con la fiabilidad necesaria, pero es necesario desarrollar una metodología capaz de analizar grandes bases de datos. Los algoritmos de Machine Learning son una buena alternativa para este tipo de problemas. Se desarrolló el algoritmo Random Boosting (**R-Boost**), que realiza de forma secuencial un muestreo aleatorio de SNP's en cada iteración sobre los cuales se aplica un predictor débil.
La población de referencia española fue usada para comparar el R-Boost frente a Bayes-A, LASSO Bayesiano y G-BLUP en términos de precisión y sesgo.
Es posible optimizar el coste de genotipado mediante la imputación de genotipos a una mayor densidad previamente a su evaluación. Se evaluó el efecto de usar genotipos de baja densidad imputados con el software *Beagle* en cuanto a su posterior habilidad predictiva. Para ello se emplearon dos métodos de evaluación R-Boost y un G-BLUP. Animales de los que se conocían los SNP's incluidos en los chips GoldenGate Bovine 3K (**3K**) y BovineLD BeadChip (**6K**), fueron imputados hasta conocer los SNP's incluidos en el BovineSNP50v2 BeadChip (**50K**). Posteriormente, un segundo proceso de imputación obtuvo los SNP's incluidos en el BovineHD BeadChip (**HD**).

RESULTADOS Y DISCUSIÓN

Las estrategias que consisten en usar como población de referencia los animales en los extremos de la distribución fenotípica obtuvieron las mejores predicciones (Figura 1). Los resultados usando 1,000 genotipos fueron 0.50 para el carácter de baja heredabilidad y 0.63 para el de heredabilidad media cuando la variable dependiente fue el fenotipo ajustado. Cuando se usaron valores genéticos como variable dependiente las correlaciones fueron 0.48 y 0.63 respectivamente. En comparación, una población de 996 machos obtuvo correlaciones de 0.48 y 0.55 en las predicciones posteriores (Jiménez-Montero et al., 2012).
Las predicciones genómicas fueron más precisas que el índice de pedigrí tradicional a la hora de predecir los resultados de futuros test de progenie como era de esperar (Tabla 1). Las ganancias en precisión debidas al empleo de la selección genómica dependen del carácter evaluado y variaron entre 0.04 (Profundidad de ubre) y 0.42 (Porcentaje de grasa) unidades de correlación de Pearson. Los resultados promediados entre caracteres mostraron que el LASSO Bayesiano obtuvo mayores correlaciones superando al R-Boost, Bayes-A y G-BLUP en 0.01, 0.03 y 0.03 unidades respectivamente. Las predicciones

obtenidas con el LASSO Bayesiano también mostraron menos desviaciones en la media, 0.02, 0.03 y 0.10 menos que Bayes-A, R-Boost y G-BLUP, respectivamente. Las predicciones usando R-Boost obtuvieron coeficientes de regresión más próximos a la unidad que el resto de métodos y los errores cuadráticos medios fueron un 2%, 10% y 12% inferiores a los obtenidos a partir del B-LASSO, Bayes-A y G-BLUP, respectivamente (González-Recio et al., 2013; Jiménez-Montero et al., 2013).

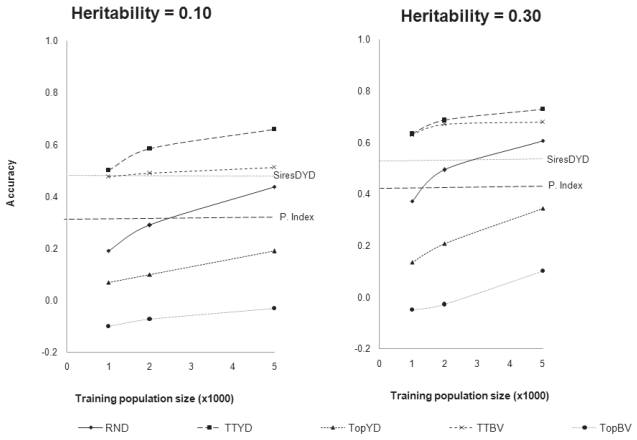


Figura 1. Precisiones estimadas para dos caracteres de heredabilidad 0.10 y 0.30 obtenidas en poblaciones de validación a partir de poblaciones de referencia de 1000, 2000 y 5000 hembras. Las estrategias de genotipado fueron: al azar (RND), mejores fenotipos (TopYD), mejores valores genéticos (TopBV), extremas en fenotipo (TTYD), extremas en valor genético (TTBV), 996 toros (SiresDYD) e índice de pedigrí tradicional.

Tabla 1. Precisión, coeficientes de regresión y error cuadrático medio (MSE) de las predicciones genómicas para cuatro metodologías de evaluación y cinco caracteres de interés económico en el vacuno lechero español.

Método	Leche (MY)	Grasa (FY)	Proteína (PY)	% de Grasa (FP)	Prof. Ubre (UD)
Precisión					
Sire-PI	0.37	0.37	0.40	0.39	0.51
B- LASSO	**0.60**	**0.61**	**0.57**	0.74	0.55
Bayes-A	0.55	**0.61**	0.55	0.65	**0.56**
R-Boost	0.54	0.60	0.50	**0.81**	0.54
G-BLUP	0.58	0.59	**0.57**	0.62	0.55
Coeficientes de regresión					
B- LASSO	0.73	0.80	0.70	1.06	0.63
Bayes-A	0.58	0.78	0.67	0.79	0.69
R-Boost	**0.87**	**0.99**	**0.80**	1.19	**0.82**
G- BLUP	0.71	0.80	0.70	**1.02**	0.64
MSE					
B- LASSO	**239992**	398	**206**	**0.03**	0.95
Bayes-A	289122	404	215	0.04	**0.90**
R-Boost	247593	**395**	216	**0.03**	0.92
G- BLUP	269619	423	219	0.04	0.95

En negrita: Método que obtiene los mejores resultados para cada criterio considerado.

Tras imputar desde 3K y 6K, se obtuvo similar capacidad predictiva a la obtenida empleando los originales en densidad 50K (Tabla 2). Sin embargo, al imputar a HD se obtuvo una pequeña mejora en la selección de individuos (Jiménez-Montero et al., Enviado).

Tabla 2. Ratio de animales correctamente clasificados de acuerdo a su ranking, conteniendo cada clase el 20% de los animales. Resultados para genotipos imputados a 50K y HD (700K) a partir de densidades 3K , 6K y 50K.

	3K50K	6K50K	50K50K	3KHD	6KHD	50KHD
Leche (MY)	0.33	0.32	**0.34**	**0.34**	0.32	**0.34**
% Grasa (FP)	0.43	0.43	0.45	0.44	0.43	**0.46**
Células somáticas (SCC)	0.32	0.33	0.34	0.32	**0.35**	0.32
Días Abiertos (DO)	0.22	0.26	0.23	0.22	0.25	**0.27**

En negrita: Genotipado con un mayor número de animales correctamente clasificados.

CONCLUSIONES

La estrategia de genotipado de hembras que proporciona la mayor correlación es la que incluye las hembras de ambas colas de la distribución de fenotipos. Mientras que la mera inclusión de las hembras élite que son las habitualmente genotipadas en las poblaciones reales produce resultados no satisfactorios en la predicción de valores genómicos.

El R-Boost es una metodología aplicable a selección genómica y competitiva en términos de capacidad predictiva.

El genotipado a baja densidad y la posterior imputación a 50K obtiene permite tomar decisiones de selección apropiadas reduciendo el coste de genotipado. Además, la capacidad predictiva para algunos caracteres puede mejorar imputando la población de referencia a HD si se emplean métodos de evaluación capaces de adaptarse a las distintas arquitecturas genéticas posibles.

REFERENCIAS BIBLIOGRÁFICAS

• Jimenez-Montero, J.A., González-Recio, O., Alenda, R. 2012. Animal. 6:1216-1224. • Jimenez-Montero, J.A., González-Recio, O., Alenda, R. 2013. J. Dairy Sci. 96:625-634. • González-Recio, O., Jimenez-Montero, J.A, Alenda, R. 2013. J. Dairy Sci. 96:*614-624* • Jimenez-Montero, J.A., Gianola, D., Weigel, K., Alenda, R., González-Recio, O. (submitted J. Dairy Sci.).

Agradecimientos: Conafe, Aberekin, Ascol, Genetycal, Xenetica Fontao, CDTI-P080250866 UPM, INIA-CC10-046, MAGRAMA. LCV Algete.

GENOMIC SELECTION IN SPANISH DAIRY CATTLE

ABSTRACT: This study includes initial stages for the implementation of genomic selection program in the Spanish Holstein population and has been used to obtain the official genomic breeding values and implement the imputation of genotypes. The objectives were: (1) To study alternative genotyping strategies for small populations, (2) to develop and validate methods for the evaluation of large data sets of genotypes, and (3) to study the effect of imputation process in predictive ability.

The most informative strategy involved genotyping of females that exhibited upper and lower extreme values when compared with traditional pedigree index and a bull reference population. Including just top animals resulted in poor results.

R-Boost resulted a competitive marker regression methodology in terms of predictive ability when compared with Bayes-A, Bayesian LASSO and G-BLUP. Optimization of the algorithm and behavior of tuning parameters were tested over real dairy cattle data reducing the time used with the original algorithm with negligible differences in accuracy and bias.

Customized GoldenGate Bovine 3K and BovineLD BeadChip genotypes imputed to a BovineSNP50v2 BeadChip using *Beagle* achieved similar predictive ability than native genotypes. Subsequently, imputing to the BovineHD BeadChip and utilizing proper methods genomic predictions may be improved.

Keywords: Genomic Selection, Dairy Cattle, Imputation