

# EVALUACIÓN PRELIMINAR DE PROCEDIMIENTOS DE SELECCIÓN GENÓMICA EN UNA POBLACIÓN DE GANADO OVINO LECHERO DE RAZA CHURRA

Sánchez, J.P.<sup>1</sup>, García-Gámez, E.<sup>2</sup>, Gutiérrez-Gil, B.<sup>2</sup> y Arranz, J. J.<sup>2</sup>

<sup>1</sup>Departamento de Ciencia Animal, Universidad Politécnica de Valencia.

<sup>2</sup>Departamento de Producción Animal, Universidad de León.

juansan@dca.upv.es

## INTRODUCCIÓN

Aunque se han desarrollado numerosas investigaciones con el objetivo de definir la determinación genética de los caracteres más relevantes en la producción lechera de oveja, el uso práctico de esta información en los programas de mejora genética es prácticamente nulo. La razón fundamental es la poca precisión en el posicionamiento de las mutaciones causales a lo largo del genoma, ya que éstas, en general, están en equilibrio de ligamiento con los microsatélites que, generalmente, se han usado como marcadores. Con la disponibilidad de paneles muy densos de marcadores moleculares, chips de SNPs, algunos de ellos van a estar en desequilibrio de ligamiento con las mutaciones causales y por lo tanto será factible usarlos en programas de selección. Para ello se han propuesto distintas alternativas, y parece ser que entre las más apropiadas se encuentran los métodos que se han llamado de selección genómica, en el que los candidatos a la selección se genotipan para un conjunto de marcadores, para los que previamente se ha obtenido un estima de su efecto, y aplicando estas estimas a los genotipos se predice el valor genético genómico de los candidatos.

En el presente estudio se pretende comparar un procedimiento de selección genómica en el que se lleva a cabo un filtrado de las regiones del genoma a considerar, basado en un procedimiento LDLA (utiliza de manera conjunta la información de ligamiento y de desequilibrio de ligamiento) uni-QTL scan; y otro en el que se modelan conjuntamente con un algoritmo de regularización el efecto de todos los SNPs de forma simultánea.

## MATERIAL Y MÉTODOS

Se emplearon YDs de ovejas de la raza Churra incluidas en el programa de mejora para la producción de leche de la raza. Estas YDs se computaron durante la evaluación de Junio de 2010. La precisión por la que ponderar estas YDs se computó siguiendo las fórmulas descritas por Garrick et al. (2009). El total de YDs (~ 900) se dividió en dos grupos, uno para el entrenamiento de los modelos y otro para validación. El segundo se constituyó con las YDs del 25% de los animales más jóvenes. Para todo estos animales se disponía de información del chip de 54K SNPs de Illumina, tras la edición de errores de genotipado, validando, los resultados del genotipado con la información genealógica, y descartar posiciones con una frecuencia alélica inferior al 5%, se procedió a la reconstrucción de haplotipos cromosómicos utilizando el programa DualPhase (Druet y Georges, 2009). Después del proceso de edición y de reconstrucción de haplotipos cromosómicos se retuvieron un total de 46,387 SNPs en los 26 autosomas.

Los caracteres considerados fueron producción de leche real (**Lreal**) y ordeñada (**L120**) por lactación, producción de leche por control (**MY**), producción y porcentaje de proteína por control (**PY**, **PP**), producción y porcentaje de grasa por control (**FY**, **FP**), recuento de células somáticas (**SCS**), prolificidad (**NV**) y éxito/fracaso a la inseminación artificial (**IA**). La tabla 1 muestra el número de YDs consideradas en los conjuntos de entrenamiento y validación.

Se llevaron a cabo dos tipos de análisis para implementar el procedimiento de selección genómica. En primer lugar (**SAM**) se implementó un procedimiento similar al empleado en las poblaciones de vacuno francesas (Boichart et al., 2010). Inicialmente se llevó a cabo una selección de regiones QTLs utilizando un modelo de QTL aleatorio que utiliza de manera conjunta la información de ligamiento y de desequilibrio de ligamiento (Druet et al., 2008). Después, y también sobre la población de entrenamiento se predijeron los efectos de los haplotipos definidos en las mencionadas regiones, así como los valores genéticos aditivos de todos los animales de la genealogía (población de entrenamiento y de validación). La predicción de las YDs de los animales de la población de validación será la suma de los efectos haplotípicos y su valor genético poligénico, ambos predichos con los datos del grupo de entrenamiento. Este procedimiento se repitió en tres situaciones diferentes de error tipo I (0.05, 0.01 y 0.001) del test de ratio de verosimilitud para declarar los QTLs como tales.

El otro método de análisis considerado es una red elástica (**EN**) (Friedman et al., 2010). La red elástica es un procedimiento de regularización que combina penalizaciones L1 y L2. Hay un parámetro  $\alpha$  que define el peso que recibe cada una de estas dos penalizaciones y otro  $\lambda$  que define la intensidad de la penalización.

$$\tilde{\beta}_{EN} = \arg \min \left\{ \sum_{i=1}^N (y_i - \bar{x}_i' \beta)^2 + \lambda \left( (1-\alpha) \sum_{j=1}^M \frac{\beta_j^2}{2} + \alpha \sum_{j=1}^M |\beta_j| \right) \right\}$$

En nuestro caso  $\alpha$  se mantuvo fijo en 0.9. Este es un procedimiento cercano al Lasso (L1) con el que se conseguirá para un gran número de SNPs de efecto muy pequeño que éste quede finalmente truncado a cero; si el valor de  $\alpha$  hubiese sido cercano a cero entonces la red elástica sería similar a una Ridge Regresión (L2). El valor de  $\lambda$  que mejor se comportó para cada carácter se obtuvo mediante una validación interna a la base de datos de entrenamiento. Por lo tanto en este caso la tabla de entrenamiento se volvió a subdividir usando el 75% de los datos para estimar los parámetros de la red elástica para un conjunto de valores de  $\lambda$ , el subconjunto de validación interna se empleó para determinar qué conjunto de estos parámetros, asociados a los distintos valores de  $\lambda$ , era el que mejor predecía YDs hasta ahora no consideradas. Se empleó el paquete glmnet de R (Friedman et al., 2010) para llevar a cabo los análisis. Finalmente el conjunto de parámetros considerados (efectos de SNPs) se utilizó para predecir las YDs del conjunto de validación, i.e. las YDs del 25% de animales más jóvenes.

Para ambos procedimientos la calidad de la predicción se evalúa computando la correlación, ponderada por la precisión de las YDs observadas, entre las YDs observadas en la base de datos de validación y sus predicciones.

## RESULTADOS Y DISCUSIÓN

La tabla 1 muestra la calidad de las predicciones realizadas variando el método de predicción. Para el procedimiento SAM se presentan distintos resultados en función del error tipo I asumido en la etapa de detección de QTLs. Entre paréntesis se muestra el número de regiones QTLs o SNPs que se consideran en la predicción.

Se observa que para FP, FY y SCS las predicciones utilizando información molecular nunca fueron mejores que la predicción de pedigrí. Para el resto de caracteres las mejoras en la predicción por el uso de información molecular variaron entre más de un 1000 % para NV, que pasa de una correlación de 0.01 a otra 0.14 al considerarse los genotipos de 268 SNPs; o más de un 230% para la leche ordeñada (L120), que pasa de una correlación de 0.07 en ausencia de información molecular a 0.16 cuando se considera el genotipo de 14 SNPs. El carácter para el que la ganancia de calidad en la predicción fue menor al considerar información molecular fue PY que sólo la incrementó un 12%, a pesar de que se consideró el genotipo de 366 SNPs. En el resto de caracteres las mejoras rondaron el 30-40%. Merece la pena señalar que salvo para MY e IA el método de uso de información molecular que mejores resultados ofreció fue la EN.

Estos resultados son preliminares pero ofrecen una evidencia de la posibilidad de uso de información de paneles de SNPs en ganado ovino de producción lechera. Sería de interés realizar estas validaciones cuando la cantidad de datos disponibles sea mayor, se prevé llegar a tener el genotipo de al menos 1500 hembras, utilizando algún otro método de los propuestos para implementar la selección genómica y probando otras alternativas para el particionado de la base de datos, la empleada probablemente no sea óptima, ya que todos las hembras de las que se tienen genotipo son coetáneas y en este caso el parentesco entre ambos grupos no es el que se tendría en una situación real de selección genómica.

## REFERENCIAS BIBLIOGRÁFICAS

- Boichard, D., Guillaume, F., Baur, A., Croiseau, P., Rossignol, M.N., Boscher, M.Y., Druet, T., Genestout, L., Eggen, A., Journaux, L. & Ducrocq V. 2010. Proceedings of the 9th World Congress on Genetics Applied to Livestock Production, Comunicación ID716.
- Druet, T., Fritz, S., Boussaha, M., Ben-jemaa, S., Guillaume, F., Derbala, D., Zelenika, D., Lechner, D., Charon, C., Boichard, D., Gut, I. G., Eggen, A. & Gautier, M. 2008. Genetics 178(4):2227-2235.
- Druet, T. & Georges, M. 2010. Genetics 184 (3): 789-798.
- Friedman, J., Hastie, T. &

Tibshirani R. 2010. J. Stat. Softw. 33(1): 1-22.. • Garrick, D.J., Taylor, J. F. & Fernando, R. L., 2009. Genet. Sel. Evol. 41:55.

**Agradecimientos:** Este estudio ha sido financiado por el proyecto CICyT AGL2009-07000.

**Tabla 1.** Correlaciones entre las predicciones de YD's de animales jóvenes y sus valores observados, entre paréntesis el número de SNPs o QTLs considerados.

Carácter <sup>1</sup>	N Entre.	N Vali.	Índice Ped.	SAM			EN
			P	$\rho$   $\alpha=0.05$	$\rho$   $\alpha=0.01$	$\rho$   $\alpha=0.001$	$\rho$
<b>L120</b>	670	222	0.05	0.07(111)	0.09(27)	0.00(4)	<b>0.16(14)</b>
<b>Lreal</b>	669	222	0.14	0.08(150)	0.04(50)	0.01(9)	<b>0.19(129)</b>
<b>MY</b>	668	209	0.20	0.19(137)	<b>0.28(47)</b>	0.25(7)	0.22(13)
<b>PP</b>	666	211	0.36	0.34(113)	0.23(27)	0.34(4)	<b>0.46(43)</b>
<b>PY</b>	666	211	0.13	0.03(118)	0.12(31)	0.00(4)	<b>0.15(366)</b>
<b>FP</b>	666	211	<b>0.46</b>	0.39(101)	0.31(29)	0.34(5)	0.43(84)
<b>FY</b>	666	211	<b>0.25</b>	0.14(99)	0.17(26)	0.20(4)	0.17(2)
<b>SCS</b>	666	211	<b>0.57</b>	0.47(124)	0.43(26)	0.40(3)	0.51(59)
<b>NV</b>	670	222	0.01	0.08(110)	0.04(27)	0.04(6)	<b>0.14(268)</b>
<b>IA</b>	593	182	0.06	<b>0.08(86)</b>	0.04(16)	0.00(1)	0.05(60)

<sup>1</sup>Ver definición en el texto.

### PRELIMINARY EVALUATION OF GENOMIC SELECTION PROCEDURES IN THE CHURRA DAIRY POPULATION

**ABSTRACT:** 52K SNP chip of about 900 Churra females where used to assess the performance of two procedures for considering this molecular information in genetic evaluation of different production and reproduction traits. In one hand a haplotype based MAS procedure after LDLA uni-QTL scan was used, in the other an Elastic Net (EN) was employed on SNP genotypes. Given the limited size of the data set, as well as its structure, results must be considered with caution. In general EN showed a better performance than any of the MAS procedures assessed. Prolificacy and milked milk were those traits showing the highest gain in accuracy after considering SNPs information. Fat traits and somatic cell score did not show any gain after including molecular information, for these traits an unexpected high accuracy was observed when only pedigree information was used, this could be an indication of a suboptimal partitioning procedure.

**Keywords:** Genomic Selection, MAS, Dairy Sheep