

## ¿PUEDEN REALIZARSE EVALUACIONES GENÓMICAS EN PEQUEÑAS POBLACIONES DE SELECCIÓN?

Ibáñez-Escriche\*, N. y Sánchez J.P.

IRTA, Genètica i Millora Animal, Av Rovira Roure 191, 25198 Lleida

\*Noelia.ibanez@irta.es

### INTRODUCCIÓN

La selección genómica (SG) es, actualmente, uno de los principales retos en mejora genética animal. Uno de los aspectos fundamentales para poder aplicar la SG con éxito en una determinada línea genética, es disponer de una población de referencia suficientemente amplia para poder estimar con precisión los valores genómicos de los caracteres de interés. En pequeñas poblaciones de selección, como por ejemplo el porcino ibérico, existe la dificultad de obtener una población de referencia adecuada. Una alternativa interesante para resolver este problema, en este tipo de poblaciones de selección, propuesta por Dekkers (2007), es la utilización conjunta de la información de las líneas "puras" (o estirpes) y sus cruces. Por otra parte, Misztal et al. (2009) propuso un método de evaluación genómica de un solo paso en el cual la matriz de parentesco se aumenta con las contribuciones de la matriz de relaciones genómicas. La principal ventaja de este método es la ponderación automática de la información fenotípica, de parentesco y genómica durante la predicción de los valores genéticos. El objetivo de este trabajo es evaluar, en pequeñas poblaciones de selección, el uso de información genómica proveniente de líneas puras y cruzadas usando el método de evaluación genómica de un solo paso propuesto por Misztal et al. (2009).

### MATERIAL Y MÉTODOS

**Modelos y Análisis Estadístico:** Los análisis estadísticos se realizaron usando un modelo animal multivariante donde se asumió que los datos de cada línea (A, B y F1) son 3 caracteres correlacionados. Así,  $\mathbf{y}' = [\mathbf{y}'_A, \mathbf{y}'_B, \mathbf{y}'_{F1}]$  es

$$\mathbf{y}_i = \mathbf{X}_i \mathbf{b}_i + \mathbf{Z}_i \mathbf{a}_i + \mathbf{e}_i,$$

donde el subíndice  $i$  indica los vectores/matrices para la correspondiente composición poblacional ( $i=A, B, F1$ ). Este modelo fue utilizado para realizar cuatro análisis estadísticos diferenciados por el tipo de información genotípica considerada en cada uno de ellos (tabla 1). Los componentes de varianza y los valores genéticos se estimaron y predijeron usando REML, considerando como matriz de covarianza entre los individuos la matriz  $\mathbf{H}$  (Legarra, et al., 2009) que combina la matriz genómica y la matriz de relaciones de parentesco aditivo. La matriz genómica se construyó siguiendo la definición de VanRaden (2008),  $\mathbf{G} = \mathbf{Z}\mathbf{Z}'/2\sum \mathbf{p}_i(1 - \mathbf{p}_i)$ ,  $\mathbf{Z}$  se define como  $\mathbf{M} - \mathbf{P}$ , donde  $\mathbf{M}$  es la matriz de los alelos de los marcadores codificándolos como el número de copias que tiene un SNP de un alelo, que en este caso puede ser 0, 1, ó 2 y  $\mathbf{P}$  es igual a dos veces la frecuencia alélica de cada locus. En la construcción de la matriz genómica se asumieron dos situaciones en relación a las frecuencia alélicas a considerar, 1) la estima de frecuencia alélica del segundo alelo (freqmix) de las poblaciones combinadas y 2) una frecuencia constante de 0.5 (freq0.5) para todos los loci.

**Simulación:** Se consideraron dos situaciones diferentes, con similar estructura. En la primera se asumió que las poblaciones A y B provenían de un origen común ( $N_e=100$ ), pero estaban separadas por 500 generaciones de apareamiento aleatorio. En la segunda situación se consideraron poblaciones con orígenes diferentes pero con el mismo tamaño efectivo ( $N_e=100$ ). En todas las simulaciones se utilizó una distribución *gamma* para muestrear los efectos de los QTL bialélicos, y una distribución *de Bernoulli* con frecuencia 0,5 para sus alelos. En todas las generaciones simuladas la tasa de mutación considerada fue de 0,000025 y la distribución de los puntos de recombinación a lo largo del genoma fue

modelada usando una función binomial de mapeo. Los fenotipos considerados en el análisis incluyeron las 3 primeras generaciones de las poblaciones A, B (200 individuos por población y generación) y la población F1 (1200 individuos) resultante de cruzar la 3ª generación de las poblaciones A y B. La genealogía incluyó 800 individuos de las poblaciones A y B (4 generaciones) y 1200 individuos de la F1. Para la generación de estos fenotipos se asumieron 100 QTLs segregando y una heredabilidad de 0.3 en la F1. La información genotípica se obtuvo a partir de 3000 marcadores SNP distribuidos en un genoma de 200 cM y formado por dos cromosomas. Durante los análisis únicamente se incluyeron los genotipos de 600 animales, pero la composición de esta información genotípica varió dependiendo del escenario a considerar (tabla1). La validación de los distintos escenarios se hizo considerando la precisión y el sesgo de los GEBVs. Para ello se calcularon la correlación y el error cuadrático entre el verdadero valor genético y el predicho para los animales de la cuarta generación de la línea A.

## RESULTADOS Y DISCUSIÓN

Las tablas 2 y 3 muestran, respectivamente, las medias y los errores estándar obtenidos con 25 réplicas para la correlación y el error cuadrático (EC), considerados como una medida de precisión y de sesgo, entre el valor verdadero y el estimado. Los resultados obtenidos muestran un aumento de correlación y una disminución de EC de las predicciones de los valores genéticos de los animales genotipados cuando se tiene en cuenta la información de los marcadores. Sin embargo, en las evaluaciones genéticas que utilizaron información genotípica, no se observaron diferencias apreciables de correlación. Este no es el caso del EC, ya que claramente la evaluación genética que combinó información de marcadores de las poblaciones A y B y su cruce (F1) fue la que obtuvo el menor valor. En relación a las frecuencias alélicas consideradas para construir la matriz **G** no se encontraron diferencias en términos de precisión ni de sesgo, estos resultados coinciden con los obtenidos por Simeone et al. (2012) con datos reales de pollos donde se combina la información genotípica, de parentesco y fenotípica de dos líneas diferentes. Entre las dos situaciones consideradas las diferencias fueron pequeñas, únicamente señalar una ligera reducción de la correlación, acompañada también de una reducción del EC, en la situación 2, cuando las líneas tienen un origen independiente. Adicionalmente, los análisis con información genómica mostraron una sobreestimación de las varianzas aditivas (resultados no mostrados) lo que coincide con lo encontrado por otros autores (Forni, et al., 2011). Los resultados obtenidos en este trabajo indicarían que es posible combinar información genómica de cruzados y de las líneas de que proceden, para mejorar los resultados de la SG en poblaciones pequeñas. No obstante, es necesario un estudio más profundo sobre la construcción de la matriz **G** teniendo en cuenta, en este tipo de evaluaciones genómicas, las frecuencias alélicas de cada población para poder mejorar así la precisión y reducir el sesgo de las varianzas genéticas y los valores genéticos.

## REFERENCIAS BIBLIOGRÁFICAS

- Dekkers J.C.M 2007. J. Anim. Sci., 85, 2104-2114.
- Forni S., Aguilar I., Misztal I 2011. Genet Sel Evol. 43(1): 1.
- Legarra A., Aguilar I., Misztal I. 2009. J. Dairy Sci., 92, 4656-4663
- Misztal I., Legarra A., Aguilar I. 2009 J. Dairy Sci., 92,4648-4655.
- Simeone et al. 2012. J. Anim. Breed. Genet. 129 3-10
- Van Raden P.M. 2008. J. Dairy Sci., 91, 4414-4423.

**Agradecimientos:** A José Luis Noguera por su inestimable aportación de ideas y su ayuda en la revisión y elaboración de este trabajo.

**Tabla 1.** Distribución del número de animales genotipados en función de la población y de la evaluación genética.

Evaluación	Población		
	A	B	F1

<b>genética*</b>			
<b>C</b>	-	-	-
<b>GAB</b>	400 (100 por generación)	400 (100 por generación)	-
<b>GF1</b>	100 (4ª generación)	100 (4ª generación)	600
<b>GABF1</b>	200 (1ª-3ª generación)	200 (1ª-3ª generación)	200
	100 (4ª generación)	100 (4ª generación)	

C: Evaluación genética sin datos de genotipos; GABC: Evaluación genética con datos de genotipos de la líneas puras; GF1: Evaluación genética con datos de genotipos de la F1 y de la población de validación; GABF1: Evaluación genética con datos de genotipos de las líneas puras y de la F1.

**Tabla 2.** Media y error estándar de la correlación, para 25 réplicas, entre los valores genéticos estimados y verdaderos de la población de validación genotipada A.

<b>Situación</b>		<b>C*</b>	<b>GABC*</b>	<b>GF1*</b>	<b>GABF1*</b>
<b>1</b>	Freqmix	0.54 (0.02)	0.69 (0.02)	0.71 (0.02)	0.72 (0.01)
	Freq0.5	0.54 (0.02)	0.69(0.01)	0.70 (0.01)	0.71 (0.01)
<b>2</b>	Freqmix	0.50 (0.02)	0.68 (0.02)	0.67 (0.02)	0.70 (0.01)
	Freq0.5	0.50 (0.02)	0.68 (0.02)	0.67 (0.02)	0.70 (0.01)

\*C: Evaluación genética sin datos de genotipos; GABC: Evaluación genética con datos de genotipos de la líneas puras; GF1: Evaluación genética con datos de genotipos de la F1 y de la población de validación; GABF1: Evaluación genética con datos de genotipos de las líneas puras y de la F1.

**Tabla 3.** Media y error estándar del error cuadrático, para 25 réplicas, entre los valores genéticos estimados y verdaderos de la población de validación genotipada A.

<b>Situación</b>		<b>C*</b>	<b>GABC*</b>	<b>GF1*</b>	<b>GABF1*</b>
<b>1</b>	Freqmix	0.78 (0.13)	0.73 (0.14)	0.75 (0.08)	0.45 (0.06)
	Freq0.5	0.78 (0.13)	0.71 (0.14)	0.75 (0.08)	0.40 (0.06)
<b>2</b>	Freqmix	0.70 (0.09)	0.61 (0.10)	0.68 (0.11)	0.35 (0.03)
	Freq0.5	0.70 (0.09)	0.62 (0.10)	0.68 (0.11)	0.39 (0.04)

\*C: Evaluación genética sin datos de genotipos; GABC: Evaluación genética con datos de genotipos de la líneas puras; GF1: Evaluación genética con datos de genotipos de la F1 y de la población de validación; GABF1: Evaluación genética con datos de genotipos de las líneas puras y de la F1.

### **¿COULD BE THE SINGLE STEP GENOMIC EVALUATION USED IN SMALL POPULATIONS OF SELECTION?**

**ABSTRACT:** The aim of this study is to evaluate under simulation the use of genomic information from purebred and crossbred lines using the Genomic Evaluation Single Step method in small populations of selection. Two scenarios with two purebred lines (A and B) and a F1 cross were simulated: 1) two purebred lines have a common origin but they are 500 generations apart and 2) two breeds without common origin. Trait phenotypic values controlled by 100 segregating QTL and with a heritability of 0.30 were simulated for the F1 individuals. Further, 3000 segregating markers on two chromosomes of 1 Morgan were chosen for analysis. The same amount of phenotypes (600 A, 600 B, 1200 F1), pedigree (800 A, 800 B, 1200 F1) and genotypes (600) data were used in all genomic evaluations. However, three different combinations of genotypes were used: only purebred, only F1 and a mixing of purebred and F1 individuals, respectively. Results did not showed a decrease of accuracy when the F1 genotype information was incorporated in genomic evaluation. Moreover, the smallest quadratic error was obtained when purebreds and F1 genotypes combination were used. These results would strengthen of using genotype crossbred information to genomic evaluate small populations with the Single Step method.

**Keywords:** Genomic selection, Crossbred data, Single Step approach, Small populations of selection