

## **EFFECTO DEL MODO DE DIVERGENCIA EVOLUTIVA EN LA PRECISIÓN DE LA SELECCIÓN GENÓMICA EN POBLACIONES MÚLTIPLES**

Mouresan, E. F., González, A., Moreno, C., Altarriba, J., Varona, L.

Unidad de Genética Cuantitativa y Mejora Animal. Universidad de Zaragoza. 50013.

Zaragoza

[e.f.mouresan@hotmail.com](mailto:e.f.mouresan@hotmail.com)

### **INTRODUCCION**

La selección genómica (Meuwissen et al., 2001) es un procedimiento de selección asistida por marcadores que estima los efectos asociados a éstos en una población de aprendizaje para predecir el mérito genético de los futuros candidatos a la selección. La precisión de la selección genómica depende del grado de desequilibrio de ligamiento entre los marcadores y las mutaciones causales de la variabilidad genética, así como de la cantidad de información fenotípica disponible para estimar los efectos asociados a los marcadores (Hayes et al., 2009). En vacuno de leche, la eficiencia de la selección genómica para incrementar la respuesta a la selección ha sido claramente demostrada (Hayes et al., 2009). Este hecho se produce gracias a la existencia de una población de referencia internacional y al uso extensivo de la inseminación artificial.

Sin embargo, la eficiencia en vacuno de carne es más cuestionable, debido a que existen poblaciones de tamaño reducido. La ausencia de una población de referencia de suficiente magnitud impide obtener estimadores de alta precisión (Seidel, 2009) en población única. Por lo tanto, la evaluación a partir de información procedente de varias poblaciones puede ser de gran utilidad. Sin embargo, la heterogeneidad de origen e historia evolutiva de las poblaciones, genera variabilidad entre ellas en cuanto a la determinación genética para los caracteres de interés zootécnico y los patrones de desequilibrio de ligamiento (de Roos et al., 2009). Como consecuencia, las estimaciones obtenidas de los efectos asociados a los marcadores en una población no son directamente extrapolables de una población a otra (Harris et al., 2008).

El objetivo de este trabajo es comparar la precisión de los valores mejorantes predichos, utilizando información fenotípica de varias poblaciones y metapoblaciones en la estimación de los efectos asociados a los marcadores, bajo distintos escenarios de densidad de marcadores, generaciones de divergencia y causas del aislamiento reproductivo.

### **MATERIAL Y MÉTODOS**

Se simuló una población base de 100 individuos (50 machos y 50 hembras). Para cada uno de ellos se simularon 30000 marcadores dialélicos localizados uniformemente a lo largo de 20 cromosomas de 1 Morgan cada uno. De estos marcadores, se seleccionaron 100 de ellos al azar para cada uno de los dos caracteres simulados y se les atribuyó un efecto aditivo obtenido por muestreo al azar de una distribución  $N(0,1)$ . Además, se simuló una tasa de mutación de  $6.6 \times 10^{-4}$ , con el objetivo de mantener la variabilidad genética. Los individuos fundadores se aparearon aleatoriamente durante 1,000 generaciones. La transmisión de la información genética se realizó mediante *gene dropping*. A partir de este momento, la población se subdividió en 4 poblaciones, también de 100 individuos cada una de ellas. En ellas, se simularon valores fenotípicos para los dos caracteres asumiendo heredabilidades 0.2 y 0.4. Las poblaciones A y B fueron sometidas a un proceso de selección fenotípica para cada uno de los dos caracteres durante 50 generaciones. Además, las poblaciones C y D se mantuvieron bajo apareamiento aleatorio también durante 50 generaciones. El proceso de obtención de estas poblaciones se ilustra en la Figura 1.

En la 50ª generación las poblaciones se expandieron a 2,000 individuos cada una de ellas y se realizó una evaluación genómica mediante el software GS3 (Legarra et al., 2012), tomando como poblaciones de entrenamiento 11 posibles alternativas, las cuatro poblaciones puras (A,B,C y D) y 7 poblaciones compuestas (A+B, A+C, A+D, B+C, B+D, C+D, A+B+C+D) con igual porcentaje de cada una de las poblaciones de origen. Para la evaluación, se utilizaron los modelos Bayes-CPi (Habier et al., 2011) y GBLUP (Meuwissen et al., 2001). Se asumieron distribuciones uniformes para las varianzas y para  $\pi$ , donde  $\pi$  es la probabilidad de que un marcador tenga efecto. Se analizó mediante McMC a partir de una cadena de 100,000 iteraciones después de descartar las 20,000 primeras.

A partir de la estimación de los efectos asociados a los marcadores en las poblaciones de aprendizaje, se obtuvo la predicción de los valores mejorantes de los candidatos a la selección en todas las poblaciones. La precisión se estimó mediante la correlación entre los valores predichos y simulados de los valores mejorantes de los candidatos a la selección

## RESULTADOS Y DISCUSIÓN

Los resultados de precisión de los valores mejorantes predichos en las poblaciones puras y para el carácter 1 mediante BayesCPi se presentan en la Tabla 1. En ella se observa que la capacidad predictiva es superior en las poblaciones que no han sufrido un proceso de selección reciente (C y D) y en las que han sido seleccionadas para otro carácter (B), que en la población en la cual el carácter de interés ha sido objeto de selección (A). Este resultado se debe a la disminución de la variabilidad genética en la población seleccionada y la fijación de una mayor proporción de algunos de los genes causales de la variabilidad genética. Por otra parte, se constata que entre poblaciones separadas durante 50 generaciones, la capacidad predictiva de una población a otra es baja. Los resultados utilizando GBLUP fueron equivalentes aunque con una precisión inferior.

Cuando se utilizaron bases de datos de entrenamiento compuestas por individuos de dos poblaciones, la precisión de la evaluación para los individuos de las poblaciones implicadas fue en torno a 0.35 y muy inferior para las poblaciones no implicadas en la base de datos de entrenamiento (Tabla 2). Finalmente, cuando se utilizó una base de datos de entrenamiento compuesta por individuos de todas las poblaciones, los resultados obtenidos fueron mejores que a partir de 2 poblaciones para todas las poblaciones implicadas (desde 0.380 a 0.513).

Los resultados para el carácter 2, con heredabilidad 0.40, fueron equivalentes, aunque, como era esperable, la precisión fue de mayor magnitud (resultados no mostrados). En estos momentos se está trabajando en la ejecución de varias réplicas del experimento de simulación para obtener resultados de mayor precisión y calcular el grado de incertidumbre de los mismos.

## REFERENCIAS BIBLIOGRAFICAS

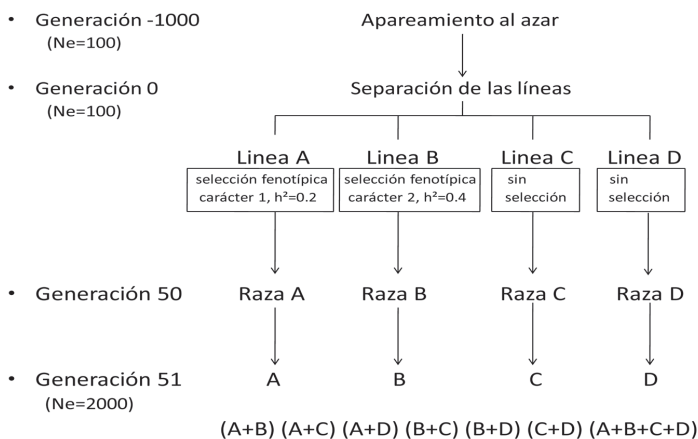
- de Roos, A. P. W., Hayes, B. J., Goddard, M. E. 2009. *Genetics* 183: 1545-1553.
- Habier, D., Fernando, R., Kizilkaya, K., Garrick, D., 2011. *BMC Bioinformatics* 12, 186.
- Harris, B. L., Johnson, D. L., Spelman, R.J. 2008. *Proc. ICAR 36<sup>th</sup> Session*, pp. 325-330.
- Hayes, B., Bowman, P., Chamberlain, A., Verbyla, K., Goddard, M. 2009. *Genetics Selection Evolution* 41, 51.
- Hayes, B. J., Bowman, P. J., Chamberlain, A. J., Goddard, M. E. 2009. *J. Dairy Sci.* 92, 433-443.
- Legarra, A., Ricard, A., Filangi, O. 2012. <http://snp.toulouse.inra.fr/~alegarra/>
- Meuwissen, T. H. E., Hayes, B. J., Goddard, M. E. 2001. *Genetics* 157, 1819-1829.
- Seider, G. E., Jr. 2009. *Reproduction, Fertility and Development* 22: 138-144.

**Tabla 1.** Precisión de los valores mejorantes predichos en población pura para el carácter 1.

Poblaciones de aprendizaje				
Validación	A	B	C	D
A	0.629	0.091	0.124	0.074
B	0.016	0.690	0.075	0.275
C	0.161	0.096	0.689	0.172
D	0.114	0.086	0.059	0.725

**Tabla 2.** Precisión de los valores mejorantes predichos en poblaciones compuestas para el carácter 1.

Poblaciones de aprendizaje							
Validación	A+B	A+C	A+D	B+C	B+D	C+D	A+B+C+D
A	0.312	0.276	0.319	0.049	0.037	0.027	0.380
B	0.289	0.056	0.043	0.383	0.351	0.052	0.513
C	0.015	0.331	0.054	0.318	0.052	0.314	0.392
D	0.106	0.037	0.386	0.103	0.346	0.360	0.453



**Figura 1.** Representación esquemática de las poblaciones simuladas.

## THE EFFECT OF THE TYPE OF THE EVOLUTIONARY DIVERGENCE ON THE ACCURACY OF GENOMIC SELECTION IN MULTIPLE POPULATIONS

**ABSTRACT:** In beef cattle, the small size of the population of each breed does not allow us to have a training set of sufficient size as to achieve high accuracies of genomic selection, and the evaluations in one breed do not apply to others. Using admixed populations for training might give a solution to these problems. The results of this study show that training in purebred populations gives good accuracies when validating in the same population, although slightly lower when the population is under selection for the trait of interest. Training in admixed populations gives higher accuracies for the breeds that participate in the training set than for the ones that do not, but these accuracies are significantly lower than those obtained from training in purebred population. Finally, when the training set is made of all 4 breeds, the accuracies are slightly better for all breeds than those obtained with the 2-breed admixed population.

**Keywords:** genomic selection, multiple populations